AI-DRIVEN IDENTIFICATION OF MELANOMA RISK FACTORS USING CHOROIDAL NEVI RETINAL IMAGES

BY

MUHAMMAD HUZAIFA KHAN SURI

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Adviser:

Assistant Professor Yogatheesan Varatharajah

ABSTRACT

This thesis explores the potential of using choroidal nevi retinal images to identify melanoma risk factors, employing advanced machine learning techniques. Utilizing a comprehensive dataset annotated by ocular oncology specialists, the study develops and validates models capable of distinguishing benign nevi from those at risk of transforming into melanoma. Our models achieve a peak Area Under the Curve (AUC) of 0.93 for identifying significant risk factors, outperforming baseline models such as ResNet-50. A significant focus of this research is on enhancing the interpretability of these AI models, ensuring that the diagnostic predictions are transparent and can be understood by clinicians. This approach not only improves trust in AI-driven diagnostics but also facilitates deeper insights into the decision-making process of the models. Moreover, the models demonstrate robust performance under various imaging conditions, including a maximum performance drop of only 5.28% at 40% zoom out, highlighting their utility in diverse clinical settings. The results demonstrate the efficacy of the models in identifying key risk factors and predicting nevi transformation, which could lead to earlier interventions and potentially improved patient outcomes in ophthalmology. This thesis sets the groundwork for future research aimed at integrating AI with traditional imaging techniques to create more robust, interpretable, and clinically applicable diagnostic tools.

To my parents, siblings, and wife, for their love and support.

ACKNOWLEDGMENTS

First and foremost, I would like to express my profound gratitude to my advisor, Professor Yogatheesan Varatharajah, whose guidance and support were instrumental in shaping this research. His insights and encouragement were pivotal at every stage of this thesis, particularly when I was uncertain about the direction of my work. His belief in my capabilities and the inspiring research idea he provided set the foundation for this project. I am also grateful to my lab colleagues, especially Neeraj Wagh, for their collaboration and support throughout this journey. Additionally, I thank the Electrical and Computer Engineering Department at the University of Illinois at Urbana-Champaign (UIUC) for the opportunity to serve as a Teaching Assistant.

My sincere appreciation extends to the team at the Mayo Clinic, who were essential to the success of this project. I am especially thankful to Dr. Raymond Iezzi, Dr. Lauren Dalvin, Connor Lentz, David Leske, and Mostafa Sadegh Mousavi for their expert guidance and support. Coming from a non-medical background, their insights into the clinical aspects and their assistance with the dataset were invaluable. I am also grateful to the Mayo Clinic for providing the grant that supported this research, enabling us to pursue our objectives without constraint.

I am thankful for the support provided by The Center for Artificial Intelligence Innovation (CAII) at the National Center for Supercomputer Applications, UIUC. Special thanks to Dr. Volodymyr Kindratenko and Shirui Luo, as well as my fellow colleagues at the center, whose assistance was crucial in facilitating my research with cutting-edge resources and technologies.

Lastly, I owe a tremendous debt of gratitude to my family — my parents, siblings, and particularly my wife — who have provided unwavering support and encouragement throughout this endeavor. Their endless love and motivation have been my strength and inspiration. Special mention to my friends for their support during this challenging yet rewarding journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AUC	Area Under the Curve
AI	Artificial Intelligence
AMD	Age-Related Macular Degeneration
ANN	Artificial Neural Network
BRVO	Branch Retinal Vein Occlusion
CNN	Convolutional Neural Network
CRVO	Central Retinal Vein Occlusion
CV	Computer Vision
DR	Diabetic Retinopathy
EDI-OCT	Enhanced Depth Imaging Optical Coherence Tomography
EHR	Electronic Health Records
FA	Fluorescein Angiography
FAF	Fundus Autofluorescence
FPR	False Positive Rate
Grad-CAM	Gradient-weighted Class Activation Mapping
ICG	Indocyanine Green Angiography
IRB	Institutional Review Board
LLM	Large Language Models
ML	Machine Learning
NLP	Natural Language Processing

OCT	Optical Coherence Tomography
ON	Optic Nerve
ROC	Receiver Operating Characteristics
SGD	Stochastic Gradient Descent
TPR	True Positive Rate
t-SNE	t-Distributed Stochastic Neighbor Embedding
ViT	Vision Transformer

CHAPTER 1

INTRODUCTION

A significant challenge confronting the global healthcare community is the insufficient availability of medical professionals to screen and diagnose specialized diseases. This deficit is especially felt in remote and developing regions, where geographic and economic barriers limit access to trained specialists. Patients in these areas are often dependent on primary care providers, resulting in referrals to specialized medical professionals and notable delays in the diagnosis and treatment processes. For individuals afflicted with terminal diseases, these delays can substantially diminish survival prospects.

Ophthalmology is an area of medicine profoundly impacted by the shortage of specialists. The number of ophthalmologists has remained constant since 1990, while the number of patients over 65 is projected to increase by 42% and 83% by 2030 and 2050, respectively [1]. A shortage of specialists poses increased risks to patients with ocular conditions such as choroidal melanoma, the most common intraocular malignancy in adults, with an incidence of approximately five cases per million people in the United States [2]. Close to 50% of patients with choroidal melanoma will develop metastasis within 10 years of diagnosis, leading to a very poor prognosis with a survival time of 6–12 months [3, 4]. The timely detection of this condition plays a pivotal role in achieving enhanced local tumor control and ultimately improving prognosis.

The precursor lesion to choroidal melanoma, called choroidal nevus, is a benign tumor with a prevalence estimated as high as 5% of the general population [5]. Though no specific guidelines exist, frequent screening with multimodal imaging is typically recommended to detect possible conversion to melanoma. However, the accessibility of advanced imaging facilities is limited in underserved regions; thus, color fundus photography is often utilized as it is the most accessible, cost-effective, and user-friendly method for documenting retinal and optic nerve conditions. Moreover, the challenge of identifying lesions that have transformed or are at high risk of transforming continues to complicate the responsibilities of eye care providers, frequently necessitating referrals to retina and ocular oncology specialists [6]. Interestingly, in a recent study, it was found that among a group of 300 neuro-ophthalmology patients, about 40% of referred patients had been misdiagnosed, while 49% patients had been at least partially misdiagnosed [7].

Artificial Intelligence (AI) holds considerable promise as a diagnostic aid in medicine, with the potential to significantly streamline the diagnostic pathway by reducing unnecessary specialist referrals. Research has indicated that certain risk factors serve as early predictors of the progression from choroidal nevi to melanomas. While some of these indicators are readily identifiable, others require detection by skilled specialists. This study proposes a framework for a diagnostic tool designed to accurately identify these risk factors, thereby supporting eye care providers in the screening process. Utilizing data annotated by ocular oncology specialists and cutting-edge AI models, this tool offers a comprehensive, interpretable, and user-friendly solution for enhancing diagnostic accuracy in ophthalmological care.

The following diagram summarizes the approach utilized in this study to achieve the desired objectives:

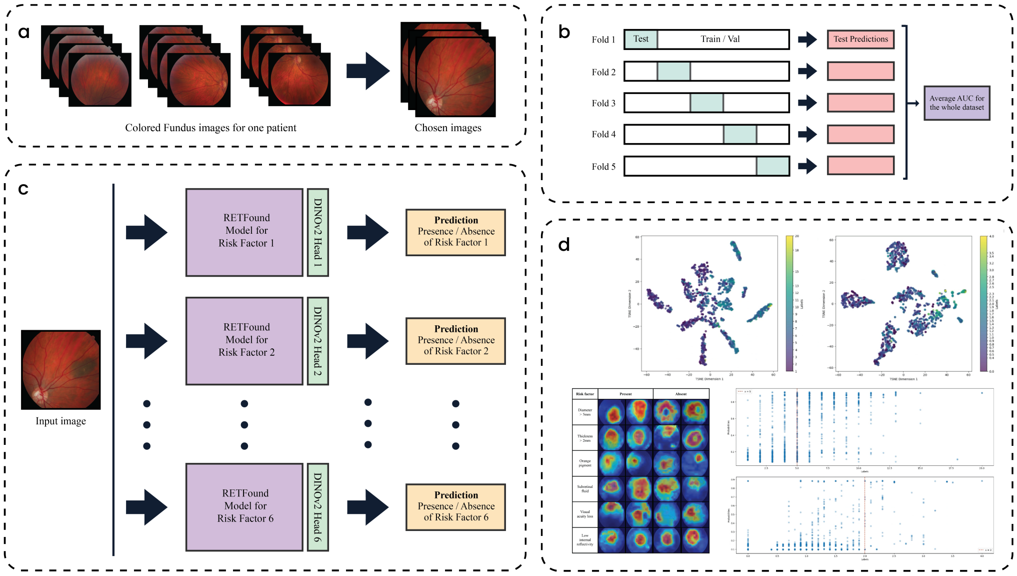

Figure 1.1: Proposed framework for this study: (a) data preparation, (b) training strategy, (c) model outcomes, and (d) interpretability

CHAPTER 2

BACKGROUND

2.1 AI in Healthcare [1]

The advent of artificial intelligence (AI) in healthcare marks a significant shift towards more precise, efficient, and patient-centered medical services. AI's ability to process vast amounts of data rapidly and with high accuracy is revolutionizing various aspects of the healthcare industry, from diagnostic procedures to patient management and drug discovery. This transformation is facilitated by advancements in machine learning (ML), computer vision (CV), and natural language processing (NLP), enabling healthcare systems to offer enhanced care delivery.

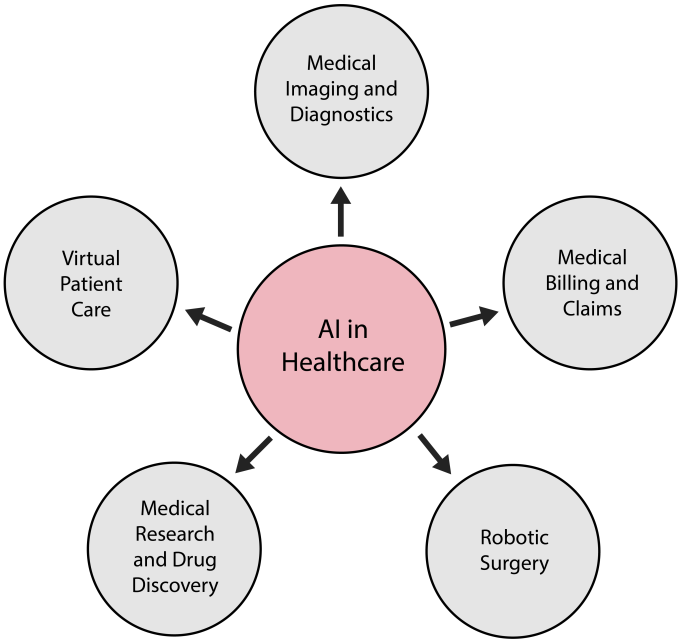

Figure 2.1: AI for healthcare, summarized

---

[1] Adapted from [8] and [9]

2.1.1 AI in Medical Imaging and Diagnostics

One of the most prominent applications of AI in healthcare is in the field of medical imaging and diagnostics. AI algorithms, particularly those based on deep learning, are increasingly employed to analyze images from X-rays, MRIs, and CT scans to detect and diagnose conditions from orthopedic issues to oncological diseases. For example, AI systems have been developed to identify patterns in imaging data that are indicative of diseases such as breast cancer, lung nodules, and cardiovascular abnormalities. These systems not only improve the accuracy of diagnoses but also significantly reduce the time taken to interpret scans, thus speeding up the patient care process.

2.1.2 AI in Virtual Patient Care

AI significantly enhances virtual patient care by integrating with Electronic Health Records (EHRs) to facilitate the management of chronic diseases and continuous patient monitoring. Through the use of ML and NLP, AI analyzes data collected via wearable technologies and EHRs to provide real-time feedback and health alerts to both patients and healthcare providers. This integration allows for a comprehensive view of a patient's historical and real-time health data, enhancing decision-making and enabling a proactive approach to healthcare. The advent of Large Language Models (LLMs) has also significantly improved virtual assistance by understanding and responding to patient queries and needs [10]. Such AI-driven systems are crucial in preventing complications, reducing hospital readmissions, and improving the overall quality of life for patients with chronic conditions, thereby transforming telehealth services into more personalized and efficient care solutions.

2.1.3 AI in Medical Research and Drug Discovery

In the realm of research and drug discovery, AI has taken significant strides by enabling the analysis of complex biological data and speeding up the drug development process. AI systems use predictive models to identify potential drug candidates and simulate their interactions with biological targets. This not only reduces the time and cost associated with traditional drug discovery but also increases the likelihood of finding effective therapies. Moreover, AI-driven analysis of medical literature and clinical data helps in identifying novel therapeutic pathways and understanding disease mechanisms.

2.1.4 AI in Medical Billing and Claims

The impact of AI also extends to administrative functions such as medical billing and claims processing, where it automates routine tasks, reduces errors, and ensures compliance with healthcare regulations. AI systems streamline these processes by extracting relevant information from unstructured data, thus allowing healthcare professionals to focus more on patient care rather than administrative duties.

2.1.5 AI in Robotic Surgery

AI-enhanced robotic surgery represents a transformative development in surgical practices, significantly improving outcomes and patient experiences. In many leading hospitals, surgeons now employ robotic systems that allow them to conduct complex procedures with unprecedented precision. By manipulating robotic arms from a computer console, surgeons can gain a magnified, three-dimensional view of the surgical site, enhancing their ability to perform delicate operations. This technology not only facilitates a high degree of precision but also allows for minimally invasive procedures, which are associated with fewer complications, reduced postoperative pain, and faster recovery times. As AI continues to evolve, the integration of these technologies in robotic surgery is poised to further revolutionize surgical practices, making surgeries safer and more accessible while reducing the physical strain on surgeons and enhancing collaborative team efforts during operations [11].

2.2 Ophthalmic Imaging

Ophthalmic imaging plays a pivotal role in the diagnosis and management of intraocular tumors, offering a diverse array of techniques each suited to different aspects of tumor evaluation. Various imaging techniques are employed, each with its specific utilities, enabling precise assessment and treatment monitoring. Here, we explore the primary imaging modalities used in ophthalmology, their collection methods, and their diagnostic significance. The choice of imaging technique depends on the specific characteristics of the tumor and the clinical information needed. As technology advances, these imaging tools continually evolve, offering finer detail and broader diagnostic capabilities, which are essential for the effective management of intraocular conditions.

2.2.1 Fundus Photography and Ultrawidefield Imaging [2]

Fundus photography, utilizing specialized cameras through dilated pupils, has long been integral to ophthalmology for capturing detailed images of the retina, revealing vital information about retinal lesions' color, size, and features. Ultrawidefield (UWF) imaging enhances this capability by offering extensive visualization of peripheral retinal lesions, invaluable for assessing lesion morphology and monitoring progressive changes. This technology serves in diagnosing and managing a variety of ocular conditions beyond tumors, such as diabetic retinopathy, age-related macular degeneration, glaucoma, and retinopathy of prematurity. Figure 2.8 in Section 2.3 is a typical fundus photograph, which is the main focus of this study. Figure 2.2 is an example of a UWF image. Despite its advantages, UWF imaging has limitations, including compromised color fidelity and peripheral distortion. Advances in retinal imaging, such as non-mydriatic cameras and integration of artificial intelligence, continue to refine the accuracy and diagnostic utility of these visual assessments, aiding in the early detection and treatment of these sight-threatening conditions.

2.2.2 Ultrasonography (A-Scan and B-Scan) [4]

Ultrasonography remains a cornerstone in medical imaging, particularly effective for assessing a variety of intraocular conditions, including tumors. A-Scan ultrasonography provides quantitative data essential for detailed characterization of ocular structures, such as tumor height, internal reflectivity, and

---

[2] Adapted from [12]

[3] Image from [13]

[4] Adapted from [14]

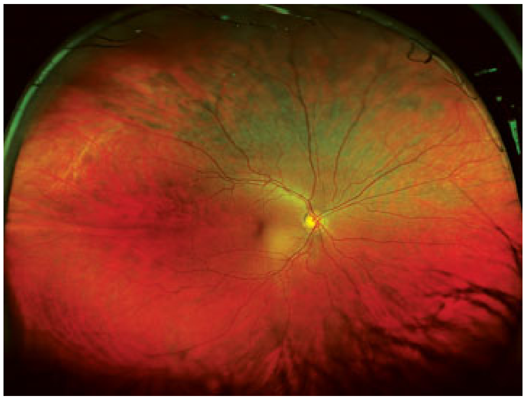

Figure 2.2: An ultrawidefield (UWF) image [3]

vascularity. Meanwhile, B-Scan ultrasonography offers a two-dimensional spatial view, depicting the shape, size, and exact location of lesions within the eye, critical for identifying potential extraocular extensions. These imaging techniques are invaluable not only for tumors but also for diagnosing and managing other significant ocular issues such as vitreous hemorrhages, retinal detachments, and ocular infections. Additionally, they are instrumental in detecting foreign bodies and evaluating ocular impacts of systemic diseases like diabetes and hypertension, which may manifest as changes in the eye's internal structure. By providing real-time, comprehensive visualizations, ultrasonography is an indispensable tool in the field of ophthalmology.

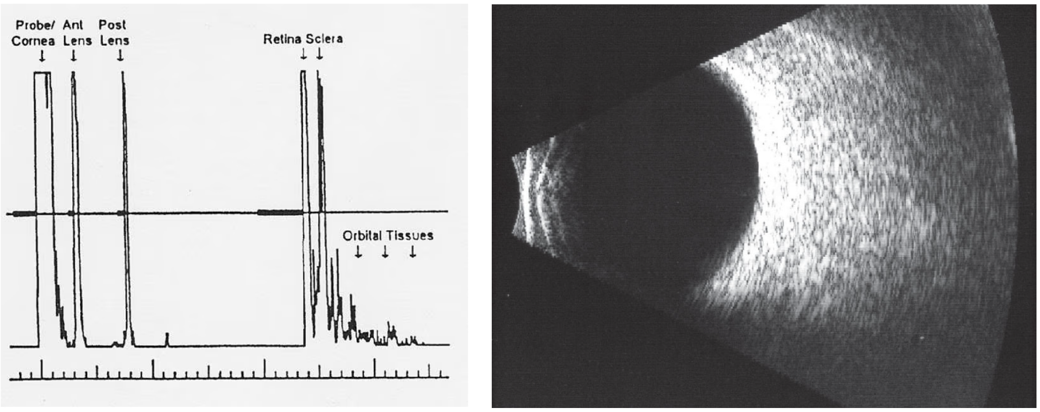

Figure 2.3: A-Scan (left) and B-Scan (right) Ultrasonography [5]

---

[5] Images from [15] and [16]

2.2.3 Fluorescein and Indocyanine Green Angiography [6]

Fluorescein angiography (FA) and Indocyanine Green (ICG) angiography are advanced diagnostic tools used primarily in ophthalmology to assess vascular structures within the retina and choroid. FA is particularly adept at identifying retinal changes such as leakage or neovascularization, which are critical for diagnosing conditions like diabetic retinopathy, macular edema, and macular degeneration. However, FA does not effectively image the choroidal circulation, which is where ICG angiography excels. ICG angiography provides enhanced visualization of the choroidal vasculature and can penetrate through hemorrhages, making it invaluable not just for assessing obscured tumors but also for conditions like central serous chorioretinopathy, polypoidal choroidal vasculopathy, and blockages of retinal veins such as branch retinal vein occlusion (BRVO) or central retinal vein occlusion (CRVO). Both these angiographic techniques are essential for tracking disease progression, targeting treatment areas, and ultimately aiding in the effective management of various ocular diseases.

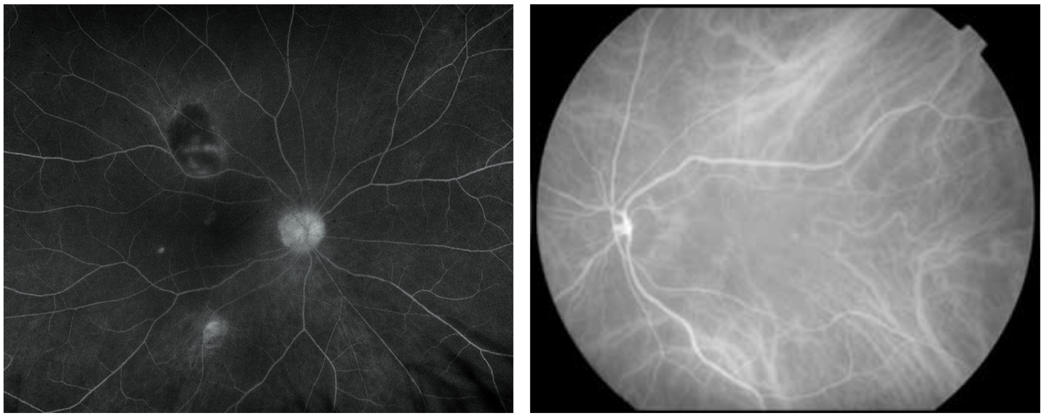

Figure 2.4: Fluorescein (left) and Indocyanine Green (right) Angiography scans [7]

2.2.4 Optical Coherence Tomography [8]

Optical Coherence Tomography (OCT) has transformed ophthalmic imaging by providing high-resolution, cross-sectional views of the eye, crucial for

---

[6] Adapted from [17]

[7] Images from [18] and [19]

[8] Adapted from [20]

diagnosing and monitoring a variety of ocular conditions. It is particularly valuable for detailed visualization of retinal layers, aiding in the assessment of intraocular tumors and deeper structures like the choroid and sclera through Enhanced Depth Imaging OCT (EDI-OCT). OCT is indispensable for managing diseases such as glaucoma, by measuring retinal and optic nerve thickness, age-related macular degeneration, diabetic retinopathy, and macular edema. It also assists in the evaluation of macular pucker and holes, offering critical insights that guide surgical planning and disease management, making it a vital tool in modern ophthalmology.

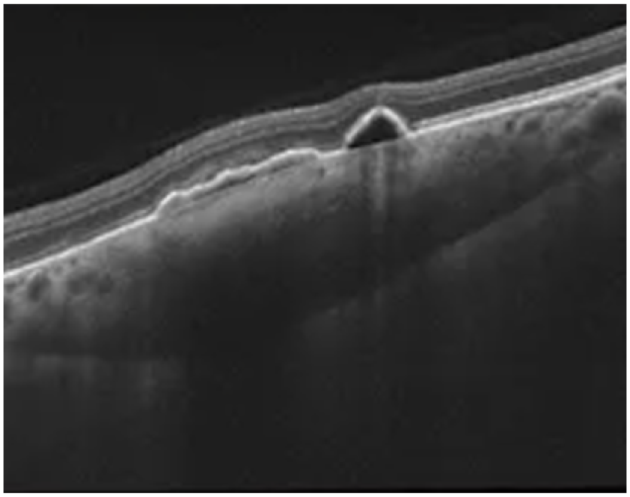

Figure 2.5: An Optical Coherence Tomography (OCT) scan [9]

2.2.5 Fundus Autofluorescence [10]

Fundus Autofluorescence (FAF) is a non-invasive imaging technique that exploits the natural emission of light from fluorophores within the eye, like lipofuscin, which accumulates due to age or cellular stress. This method is particularly useful for identifying early signs of diseases by detecting abnormal autofluorescence patterns indicative of metabolic changes. While FAF is instrumental in diagnosing ocular tumors such as melanomas and nevi, its applications extend to a variety of other retinal conditions. It is effective in tracking the progression of age-related macular degeneration (AMD), particularly geographic atrophy, and in managing retinal dystrophies like Star-

---

[9] Image from [21]

[10] Adapted from [22]

gardt disease and retinitis pigmentosa. FAF also helps in assessing central serous chorioretinopathy, hydroxychloroquine toxicity, and pattern dystrophies, making it a versatile tool in the detection and monitoring of several retinal diseases where changes in the retinal pigment epithelium play a crucial role.

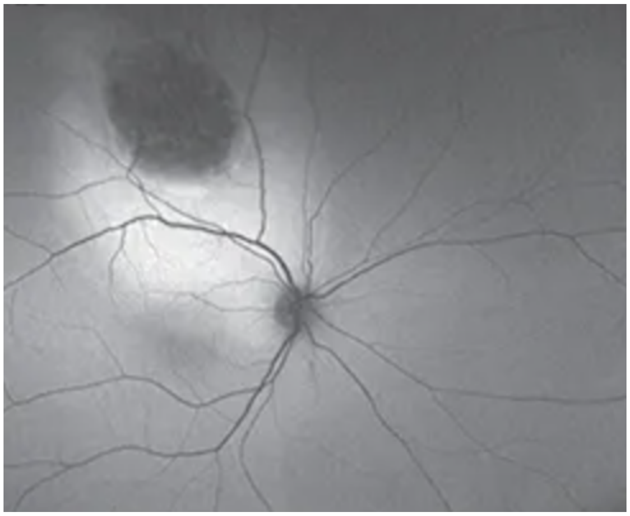

Figure 2.6: A Fundus Autofluorescence (FAF) scan [11]

2.3 Retina [12]

The retina is a vital, light-sensitive layer of tissue located at the back of the eye's interior. This complex layer is responsible for capturing light photons that enter the eye, converting them into electrical and chemical signals, and transmitting them to the brain to create visual images. The retina is a part of the eye's innermost layer, along with the vascular choroid and fibrous sclera, which together form the ocular fundus. The retina itself consists of several key structures, including the optic nerve, which transmits visual information to the brain, the macula (especially the fovea at its center), which is crucial for sharp central vision, and numerous supporting cells and blood vessels that maintain the retina's health and functionality.

---

[11] Image from [21]

[12] Adapted from [23]

[13] Image from [24]

[14] Image from [25]

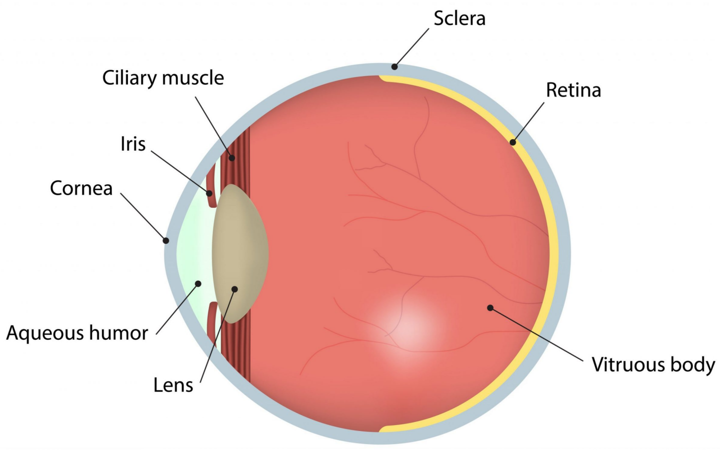

Figure 2.7: Cross-section of the human eye[13]

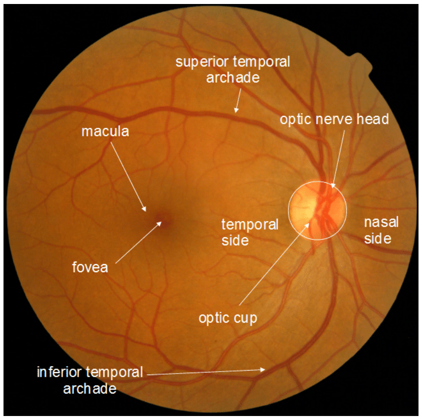

Figure 2.8: A labeled diagram of a retina[14]

Retinal health is crucial, yet several diseases can affect it, impacting vision quality and life overall. Diseases such as age-related macular degeneration (AMD), diabetic retinopathy, and retinitis pigmentosa directly affect the retina, leading to varying degrees of vision loss. AMD, for example, deteriorates the macula, leading to loss of central vision, while diabetic retinopathy involves damage to the retinal blood vessels, potentially resulting in blindness. Other conditions like glaucoma and retinal detachment also pose significant threats to vision by affecting the optic nerve and the structural integrity of the retina, respectively.

2.4 Choroidal Nevus and Melanoma [15]

Choroidal nevus and melanoma are conditions of the eye that involve the choroid, a layer of blood vessels between the retina and the sclera. A choroidal nevus is essentially a benign mole found within the eye, similar to a skin mole, and is quite common with varying prevalence but generally considered harmless. In contrast, a choroidal melanoma is a malignant tumor and is far rarer yet more dangerous. Despite the common benign nature of choroidal nevi, there is a small but significant risk that these can transform into malignant melanomas. This transformation is estimated to occur in approximately 1 in $8,845$ cases annually, with the risk slightly increasing with age. Studies have demonstrated the progression of a nevus into melanoma: a recently published longitudinal study of 3806 choroidal nevi, imaged with optical coherence tomography (OCT), ultrasonography, and standard wavelength autofluorescence, revealed transformation into melanoma in 5.8% patients at 5 years and 13.9% patients at 10 years, using Kaplan–Meier analysis [27].

Both choroidal nevi and melanomas can appear similar, making differentiation a critical aspect of ophthalmic diagnosis. Both can exhibit pigmentation and affect similar locations within the choroid, and both may present with or without symptoms such as visual disturbances or apparent visual field defects. However, choroidal melanomas tend to demonstrate more aggressive features such as irregular shapes, the presence of orange pigment, and subretinal fluid, which are less common in nevi. The progression of a nevus into a melanoma involves changes in size, thickness, and color, which necessitate careful monitoring. Early identification and treatment can drastically improve survival rates, as the potential for metastasis increases with the size and progression of the tumor.

2.5 Risk Factors Associated with Development of Nevus to Melanoma

The following risk factors have shown to be linked to the development of a choroidal nevus into melanoma by Shields et al. [27]:

---

[15] Adapted from [26]

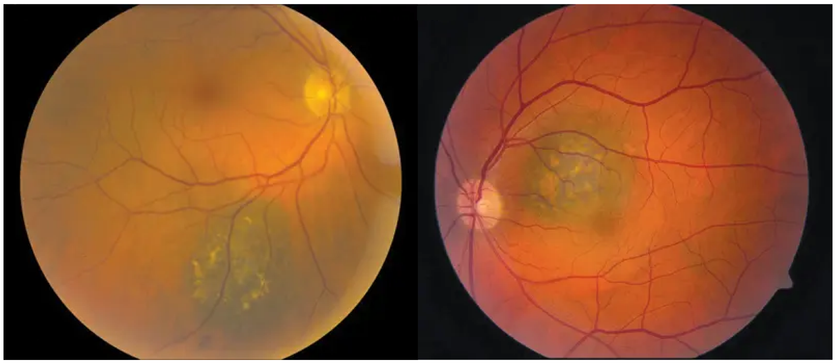

Figure 2.9: Distinguishing a nevus (left) from a melanoma (right) can be a challenging task

2.5.1 Nevus Diameter [16]

Nevus diameter is a critical metric in ophthalmology for assessing the potential risk of choroidal nevi transforming into malignant melanomas. The diameter of a nevus, when evaluated through techniques such as fundus photography or ultrasonography, provides crucial information on its growth dynamics and malignancy risk. Typically, nevi with a diameter less than 5 mm are considered less likely to evolve into melanomas; however, larger diameters, especially those over 7 mm, significantly increase the suspicion of malignancy. Regular monitoring of the diameter is essential, as a documented increase over time can be a strong indicator of malignant transformation.

2.5.2 Nevus Thickness [17]

Nevus thickness is another important diagnostic indicator in the evaluation of choroidal nevi for potential malignant transformation into choroidal melanomas. Thickness measurements, often obtained through ultrasonography or optical coherence tomography (OCT), provide essential insights into the structural characteristics of the nevus. A thickness of 2 mm or more is particularly concerning, as it significantly elevates the risk of malignancy. Regular assessments of nevus thickness are crucial for detecting early signs of transformation, as increases in thickness can suggest active cell proliferation and potentially malignant behavior.

---

[16] Adapted from [28]

[17] Adapted from [28]

2.5.3 Orange Pigment [18]

In ocular oncology, orange pigment is a biomarker often associated with active choroidal melanomas. This pigment appears as granular, orange-colored deposits primarily overlying small choroidal nevi and melanomas. Historically, its presence indicates an elevated metabolic activity within the tumor. Fundus autofluorescence (FAF) imaging techniques have proven particularly effective for detecting these bright, focal accumulations, which shine with a hyperautofluorescent glow indicative of high lipofuscin content.

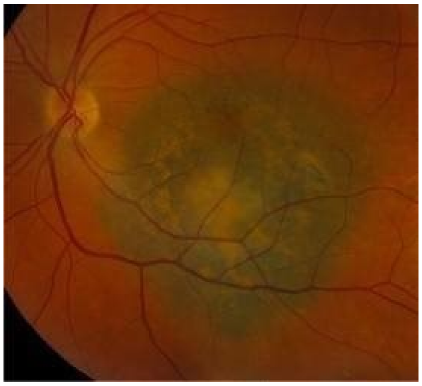

Figure 2.10: Large clumps of orange pigment in a choroidal melanoma [19]

The trajectory of orange pigment in response to therapeutic interventions, however, has been less understood and sporadically documented. Following treatments like plaque radiotherapy, changes in the visibility and intensity of orange pigment can serve as a barometer for treatment efficacy. For instance, a gradual fading of this pigment, as seen in some clinical cases, suggests a reduction in tumor activity and cellular decay. This pigment reduction is possibly due to the death of lipofuscin-containing macrophages and the overall reduction of the tumor's metabolic processes.

2.5.4 Subretinal Fluid [20]

Subretinal fluid (SRF) refers to the accumulation of fluid beneath the retina (Figure 2.11), a condition often observed following scleral buckle surgery for treating retinal detachments. This fluid, sometimes detectable only through

---

[18] Adapted from [29]

[19] Image from [30]

[20] Adapted from [31]

optical coherence tomography (OCT), can lead to varying degrees of visual recovery issues. From a technical standpoint, the persistence of SRF is crucial as it suggests ongoing retinal detachment or incomplete attachment, influencing the effectiveness of surgical interventions. Research shows that factors like the surgical method employed and the initial state of the macula can significantly impact the likelihood of persistent SRF.

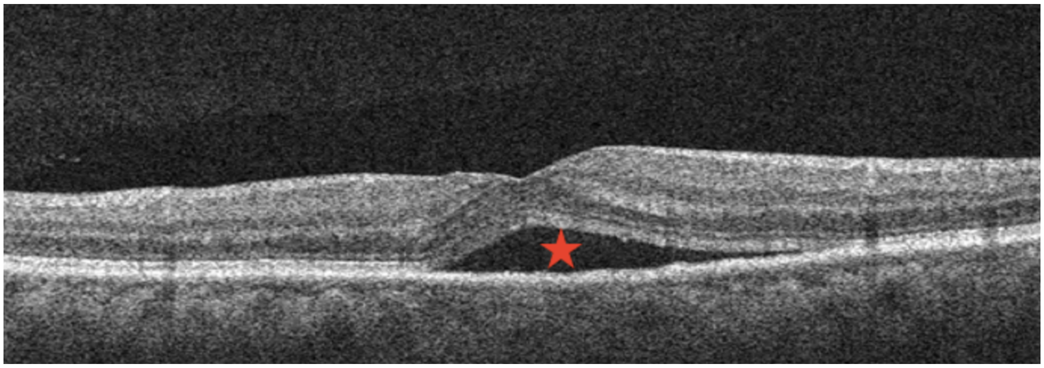

Figure 2.11: Subretinal fluid accumulation beneath the retina[21]

2.5.5 Visual Acuity [22]

Visual acuity is a term used to describe the clarity or sharpness of one's vision, which is commonly measured using the standard 20/20 vision test. This test involves reading letters from a Snellen chart (Figure 2.12) positioned 20 feet away, where "20/20" indicates that an individual can see clearly at 20 feet what should normally be seen at that distance. A variation from this norm, such as "20/100," would mean that the person must be as close as 20 feet to see what someone with standard vision sees at 100 feet. While 20/20 vision represents a benchmark for visual sharpness, it does not provide insights into other essential visual functions such as peripheral awareness, eye coordination, depth perception, focusing ability, and color vision.

Different visual conditions, such as myopia (nearsightedness), hyperopia (farsightedness), and presbyopia (age-related loss of close focusing ability), can impact visual acuity. These conditions are usually diagnosed during a comprehensive eye examination and can often be corrected with glasses,

---

[21] Image from EyeCarePD

[22] Adapted from [32]

[23] Image from [33]

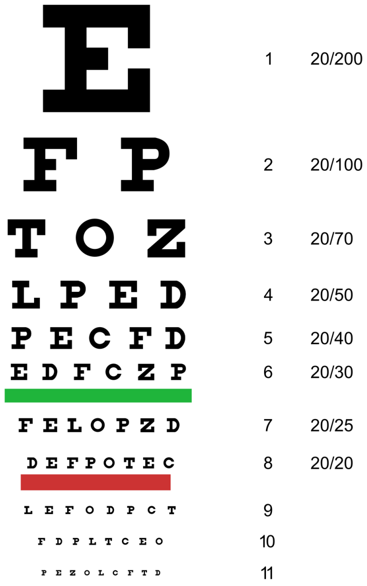

Figure 2.12: A Snellen chart[23]

contact lenses, or vision therapy. In instances where vision impairment stems from an eye disease, specific treatments such as ocular medication might be necessary.

2.5.6 Internal Reflectivity [24]

Internal reflectivity is a key parameter in ocular imaging that helps distinguish benign choroidal nevi from malignant choroidal melanomas. In the context of ocular oncology, internal reflectivity relates closely to the concept of acoustic hollowness, which is a characteristic often observed in malignant lesions such as choroidal melanomas. This property is typically assessed using B-scan ultrasonography: low to medium internal reflectivity on a B-Scan is indicative of acoustic hollowness, which contrasts with the higher reflectivity observed in benign conditions like choroidal nevi.

The measurement of internal reflectivity is crucial in early cancer detection, specifically for identifying choroidal melanoma before it manifests more visible symptoms or signs of growth. Features such as a choroidal excavation or an irregular lesion shape with low internal reflectivity are suggestive of a

---

[24] Adapted from [28]

[25] Image from [34]

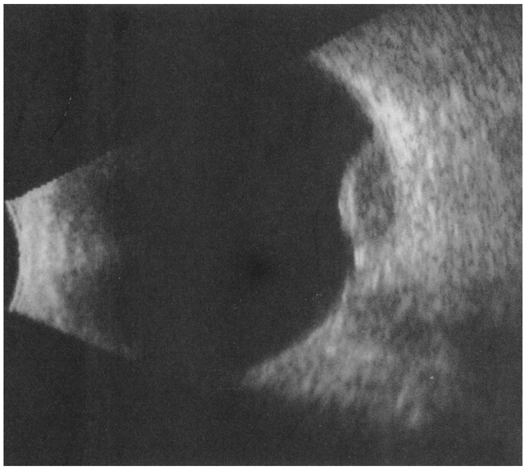

Figure 2.13: B-Scan ultrasonogram showing dome-shaped choroidal mass with acoustic hollowness, characteristic of choroidal melanoma [25]

melanoma rather than a nevus. Regular echographic follow-ups are recommended, especially when a lesion presents ambiguous features, to monitor any changes over time that might indicate malignancy.

2.6 Artificial Neural Networks [26]

Artificial neural networks (ANNs) are computational models inspired by the brain's neural networks. They mimic the way neurons in the brain communicate through a network of interconnected nodes, representing artificial neurons. The structure of an ANN includes multiple layers: the input layer, one or more hidden layers, and the output layer. Each layer contains nodes or neurons, and each neuron in one layer connects to every neuron in the next layer, forming a dense network. The input layer receives the initial data, while the hidden layers perform computations using activated functions to transform the input into something the output layer can use. Notably, deeper layers (more hidden layers) allow the network to learn more complex patterns through a process called deep learning.

The operation of an ANN involves two main phases: the forward pass and backpropagation (Figure 2.14). During the forward pass, input data is passed through the network, and each neuron applies a weighted sum of the inputs, followed by an activation function to introduce non-linearities into

---

[26] Adapted from [35]

the output. The true power of neural networks lies in the backpropagation algorithm; this is where the network learns from errors. The network adjusts its weights based on the error gradient of the loss function, which measures the difference between the actual output and the predicted output. This training is often conducted using stochastic gradient descent (SGD) or similar optimizers, which help minimize the loss function over batches of training data.

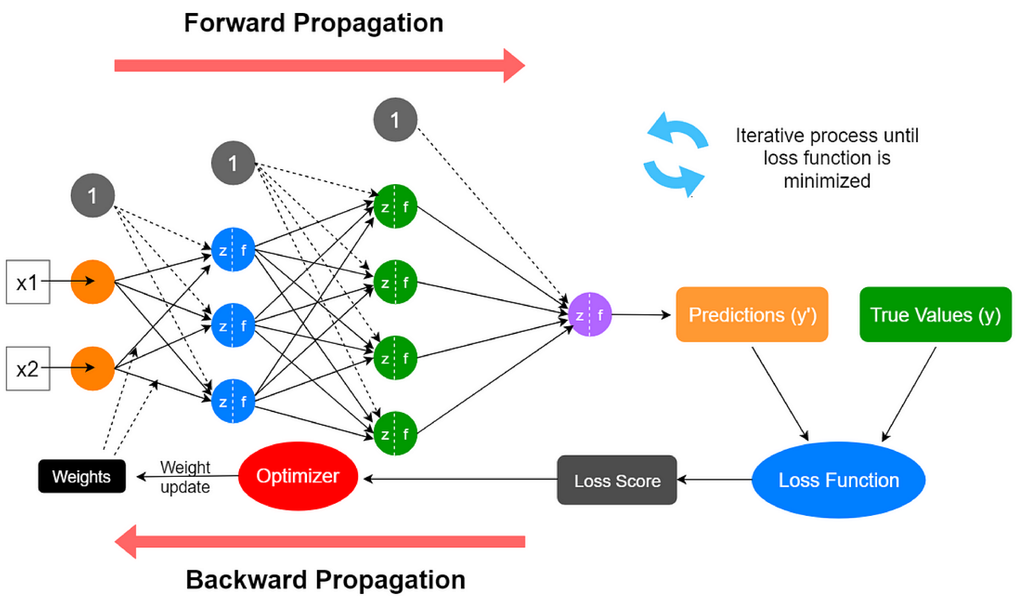

Figure 2.14: A neural network with its main phases explained[27]

The effectiveness of ANNs stems from their ability to approximate any continuous function, given sufficient data and computational power. This capability is known as the universal approximation theorem. Additionally, the depth and breadth of a network — its number of layers and neurons — can be adjusted to capture more complex patterns in the data. ANNs have shown remarkable success in tasks that involve large amounts of data and complex relationships that are difficult for humans to encode manually. Their ability to learn from examples makes them extremely versatile and powerful across a broad spectrum of applications.

---

[27] Image from [36]

2.7 Vision Transformer

Vision Transformers (ViTs) have gained prominence following the introduction of the Transformer architecture, famously detailed in the paper "Attention is All You Need" by Vaswani et al. [37], which revolutionized natural language processing with its attention mechanisms. This concept was extended to the realm of computer vision by the paper "An Image is Worth 16x16 Words" by Dosovitskiy et al. [38], which adapted the Transformer model for image classification tasks. Unlike traditional models that process images through convolutional layers, Vision Transformers treat images as sequences of fixed-sized patches (like words in a sentence), where each patch is equivalent to a "token" in NLP tasks. These patches are flattened, linearly embedded, and positionally encoded to maintain their sequence information before being fed into the transformer encoder.

The architecture of ViTs is comprised of several components that work together to process these image patches. The core component is the multi-head self-attention mechanism that allows the model to weigh the importance of different patches relative to others, enabling it to focus on the most relevant parts of an image for a given task. This is complemented by feed-forward neural networks that apply transformations to the attention outputs, and layer normalization that helps in stabilizing the training of deep networks. The entire sequence of patches passes through these layers repeatedly in what are known as transformer blocks. The effectiveness of Vision Transformers comes from their ability to learn spatial hierarchies and global dependencies between patches, which traditional convolutional neural networks (CNNs) achieve less directly. ViTs can generalize better on larger datasets and diverse tasks because they do not inherently assume any spatial hierarchies, making them flexible and scalable compared to CNNs that heavily rely on the local connectivity of image data.

2.8 Foundation Models [29]

---

[28] Image from [37]

[29] Adapted From [39]

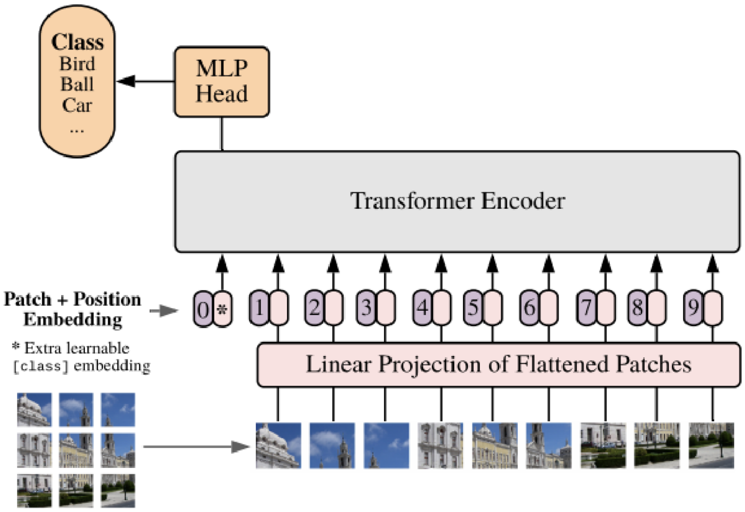

Figure 2.15: Architecture of a Vision Transformer [28]

Foundation models represent a significant shift in the landscape of artificial intelligence and machine learning. Unlike traditional machine learning models which are trained for specific tasks, foundation models are designed to provide a generalized understanding of data across a broad spectrum of tasks. These models are pretrained on vast, diverse datasets, enabling them to capture intricate patterns and structures that can later be fine-tuned to specific applications. This approach allows foundation models to serve as a versatile baseline from which specialized capabilities can be developed with additional training on more focused datasets.

The popularity of foundation models can be attributed to their unique ability to leverage transfer learning and scale, allowing them to be adapted to a wide range of industries and applications. They have revolutionized sectors such as natural language processing and computer vision. In healthcare, for instance, AI models based on attention mechanisms, a key component of many foundation models, have shown to outperform traditional convolutional models in various image classification tasks. This superiority, however, relies heavily on the availability of extensive data for training. In scenarios where data is scarce, particularly in the case of specialized diseases, the adaptive capability of foundation models becomes crucial. By starting with a model that has a broad base of pre-acquired knowledge, healthcare professionals can fine-tune algorithms to achieve high performance even with limited data, addressing specific challenges such as diagnostic imaging and personalized medicine.

2.9 Receiver Operating Characteristics Curve [30]

The Receiver Operating Characteristics (ROC) curve is a graphical representation used to evaluate the effectiveness of a binary classifier system. It plots the true positive rate (TPR) against the false positive rate (FPR) across varying threshold levels, providing insights into the classifier's performance under different conditions.

Each point on the ROC curve represents a specific balance between TPR and FPR, where TPR, also known as sensitivity, indicates the proportion of actual positives correctly identified by the classifier. In contrast, FPR, which is one minus the specificity, shows the proportion of negatives falsely identified as positives:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

Figure 2.16 shows a confusion matrix that highlights the relationship between these quantities.

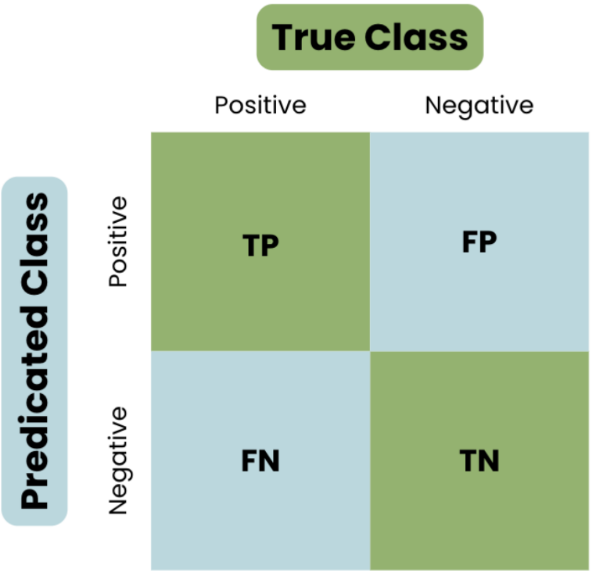

Figure 2.16: A confusion matrix [31]

---

[30] Adapted from [40]

The ROC curve typically starts at the bottom left corner (representing a classifier that predicts no positives and, hence, has zero TPR and zero FPR) and ends at the top right corner (indicating a classifier that predicts all positives).

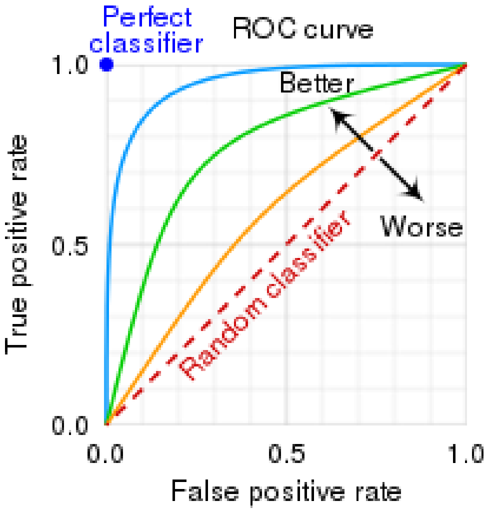

Figure 2.17: Example ROC curves [32]

The layout of an ROC curve can give significant insights into the classifier's nature, as shown in Figure 2.17. A perfect classifier would display an ROC curve that shoots straight up to the top left corner, indicating a TPR of 1 (100% sensitivity) and an FPR of 0 (100% specificity). This ideal point suggests no misclassifications. Conversely, a random classifier's performance would manifest as a diagonal line from the bottom left to the top right of the plot, reflecting a scenario where the TPR and FPR are equal — essentially no better than random guessing. The area under the ROC curve (AUC) provides a succinct scalar value summarizing the overall ability of the classifier to avoid false classification. An AUC of 1 represents a perfect classifier, while an AUC of 0.5 suggests no discriminative power better than random chance.

Through these metrics, ROC curves assist in comparing different classifiers. For instance, among multiple ROC curves, the one closest to the top left corner, or with the highest AUC, indicates the superior classifier. Such visual and quantitative analysis helps in fine-tuning the thresholds and improving the prediction accuracy of models, making ROC curves indispensable

---

[31] Image from [41]

[32] Image from [42]

in the toolkit of machine learning practitioners evaluating binary classification systems.

2.10 Grad-CAM [33]

Gradient-weighted Class Activation Mapping (Grad-CAM) is a valuable interpretability technique employed to generate visual explanations for decisions made by a wide array of convolutional neural network (CNN) based architectures. Grad-CAM operates by utilizing the gradients associated with a specific class as they propagate through the final convolutional layer. This process yields a localization map; when this map is superimposed onto the original image, regions depicted in darker shades have a greater influence on the prediction of the particular class.

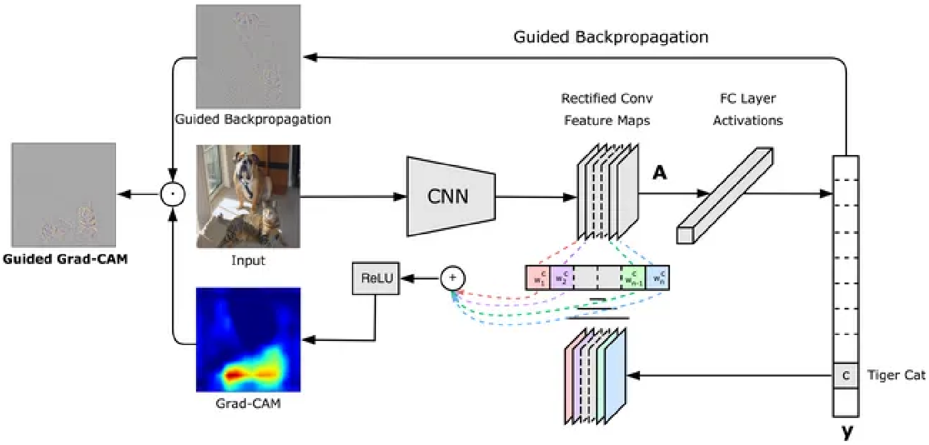

Figure 2.18: Grad-CAM visualized [34]

2.11 t-SNE [35]

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a statistical method designed to visualize complex high-dimensional data by mapping it into a two or three-dimensional space. This non-linear dimensionality reduction

---

[33] Adapted from [43]

[34] Image from [44]

[35] Adapted from [45]

technique is particularly effective in arranging data such that similar items are positioned closely in the map, while dissimilar items are spaced apart. The technique is applied across various fields such as genomics, language processing, and bioinformatics, assisting in the intuitive presentation of data clusters and patterns that are difficult to discern in high-dimensional spaces.

The operation of t-SNE involves two primary steps. Initially, it models a probability distribution over pairs of high-dimensional data points, assigning higher probabilities to similar pairs and lower probabilities to dissimilar ones. Following this, t-SNE constructs a corresponding probability distribution in the lower-dimensional space and aims to minimize the Kullback-Leibler divergence between these two distributions, preserving the local structure of data.

CHAPTER 3

RELATED WORK

Recent advancements in neural networks have demonstrated significant potential in enhancing disease detection and prognosis, often surpassing the diagnostic accuracy of highly skilled professionals [46]. In ophthalmology, deep learning techniques have been applied to fundus images to diagnose several retinal diseases. Abramoff et al. [47] and Gulshan et al. [48] were one of the first to demonstrate the use of deep learning to detect diabetic retinopathy (DR) in fundus photographs with reasonably high performance. More recently, Heydon et al. demonstrated a 95.7% sensitivity for referable DR using an assembled dataset spanning $30,000$ patients [49]. Others have utilized UWF images for DR detection [50] and OCT scans for detection of age-related macular degeneration (AMD) [51]. An AI application on a smartphone was also developed that achieved high sensitivity for detecting referable DR [52]. AI has also shown to be instrumental in detecting systemic conditions such as cardiovascular disease [53] that are linked to the retina due to hypertension and atherosclerosis. More recently, models predicting multiple retinal diseases [54] and models utilizing the newer, more powerful Vision Transformers [55, 56] have dominated performance metrics for retinal diseases.

Progress in AI is instrumental in areas like oncology, where early detection can dramatically influence treatment outcomes and patient survival rates. Kaliki et al. developed methods for detecting retinoblastoma and predicting its size using fundus photographs [3]. Similarly, Zhang et al. applied deep learning techniques to predict the presence of uveal melanoma based on iris color and iris images in a Chinese population [57]. Additionally, Yoo et al. introduced a low-shot DL model that can detect conjunctival melanoma in ocular surface images [58], further demonstrating the pivotal role of AI in enhancing diagnostic accuracies in ophthalmologic cancers.

Significant advancements have been made to differentiate between choroidal nevi and melanoma using diverse data types. Zabor et al. leveraged tabular data — comprising clinical and demographic information from patient examinations and medical records — to identify risk factors indicating the potential transformation of a nevus into melanoma [59]. They utilized a lasso regression model to distinguish between patients with nevus and melanoma, highlighting the variables with the most non-zero coefficients in their bootstrapped lasso model. Complementing this approach, Dadzie et al. explored the use of fundus photographs to differentiate nevus from melanoma, employing various color fusion techniques to enhance diagnostic accuracy [60].

A white paper by Shields et al., published in 2021, discusses the potential of AI in detecting and predicting the progression to melanoma using ocular images [61]. Thus far, only Iddir et al has described the use of an image-based machine learning algorithm to classify risk factors for choroidal lesion transformation to choroidal melanoma [62]. This study was able to accurately identify risk factors and differentiate lesions utilizing ultra-widefield fundus imaging and B-scan ultrasonography. While novel, limitations include insufficient data and a narrow focus on multimodal images that include both nevi and melanoma, thus potentially biasing the identification of crucial risk factors like subretinal fluid and orange pigment. This study underscores the necessity for a broader approach that avoids the constraints of multimodal imaging, thereby enhancing the clinical utility and adoption of these AI-driven diagnostic tools.

This study distinguishes itself by employing an extensive dataset comprising images from various cameras and lenses, focusing exclusively on extracting risk factors directly from these images. Unlike previous studies, the approach prposed in this study does not rely on multimodal imaging; while the analysis of multiple modalities may be advantageous in the accurate diagnosis of choroidal lesions, the utility of a model that exclusively utilizes fundus images is much greater in general ophthalmology or optometry settings. This streamlined method not only simplifies the diagnostic process but also increases its applicability and utility in clinical environments, potentially facilitating earlier and more accurate detections of ocular conditions.

Table 3.1: AI applications in ophthalmology

AI in the Identification and Management of Retinal Diseases

Reference	Application	Data	Approach	Pros	Cons
Abramoff et al. [47]	DR	Fundus photographs	CNNs	High performance	Limited data; no interpretability
Gulshan et al. [48]	DR	Fundus photographs	CNNs	High performance	No interpretability
Heydon et al. [49]	DR	Fundus photographs	EyeArt v2.1.0 (CNNs)	High sensitivity and specificity	No interpretability
Nagasawa et al. [50]	DR	UWF images	CNNs	High performance	Limited dataset
Natarajan et al. [52]	DR	Smartphone camera fundus photographs	Inception V3	High performance	Limited dataset; no interpretability
Lee et al. [51]	AMD	OCT scans	CNNs	High performance; interpretability	Limited dataset
Kim et al. [54]	8 retinal diseases	Fundus photographs	ResNet50, VGG19, Inception V3	High performance; interpretability	Dataset age distribution does not match disease onset age distribution
Poplin et al. [63]	Cardiovascular risk factors	Fundus photographs	Inception V3	High performance; interpretability	Limited dataset; wide confidence intervals
Wu et al. [55]	DR	Fundus photographs	ViT	High performance	Higher computational/space complexity
Xu et al. [56]	AMD	Fundus photographs	Hierarchical ViT	High performance; interpretability	Limited dataset
Kaliki et al. [64]	RB	Fundus photographs	Mobile-Net V2 SSD	High interpretability	Limited dataset; mediocre performance
Zhang et al. [57]	UM	Iris images	U-Net, Random Forest, CNN	Decent performance	Limited dataset; pre-processing susceptible to errors
Yoo et al. [58]	Conjunctival melanoma	Ocular surface images	GoogleNet, Inception V3, NASNet, ResNet50, MobileNet V2	High performance; interpretability	Limited dataset

AI in Discrimination of Choroidal Nevus from Melanoma

Reference	Application	Data	Approach	Pros	Cons
Zabor et al. [59]	Small Choroidal Melanoma vs Choroidal Nevus	Clinical features	Lasso Logistic Regression	High performance; limited interpretability	Unbalanced dataset; obtaining clinical data adds extra step
Dadzie et al. [60]	Small Choroidal Melanoma vs Choroidal Nevus	Fundus photographs	DenseNet121	High performance; interpretability	Limited dataset
Iddir et al. [62]	Risk factors of uveal melanoma in choroidal nevi	Ultra-widefield fundus imaging and B-scan ultrasonography	ResNet18	High performance + interpretability	Limited nevus images with risk factors compared to melanoma images

CHAPTER 4

DATA

This section presents a comprehensive overview of the data utilized in this study, detailing the sources, composition, and preparatory methods of the datasets employed. The research was approved by the Mayo Clinic Institutional Board Review (IRB) and was conducted in accordance with the tenets of the Declaration of Helsinki. These approvals ensured that all data collection and handling processes adhered strictly to ethical standards and legal requirements, emphasizing the protection of patient privacy and the integrity of the scientific inquiry.

4.1 Choroidal Nevus Data

This study included a retrospective detailed analysis of 18,000 fundus images and electronic health records (EHR) obtained from Mayo Clinic, Rochester, MN . The dataset included widefield color fundus photos and ultrawidefield optos photos from 509 patients, collected from July 2019 to August 2023. Each patient contributed between 2 and 40 images, captured from multiple angles of the same eye. Up to 110 attributes were obtained for each patient, including clinical (medical history, symptoms, etc.) and non-clinical attributes (age, gender, etc.). The basis of data collection was the presence of a choroidal nevus in either or both eyes. Figure 4.1 displays some examples of choroidal nevi with/without risk factors that have shown to be highly associated with the potential development of a choroidal nevus into a melanoma.

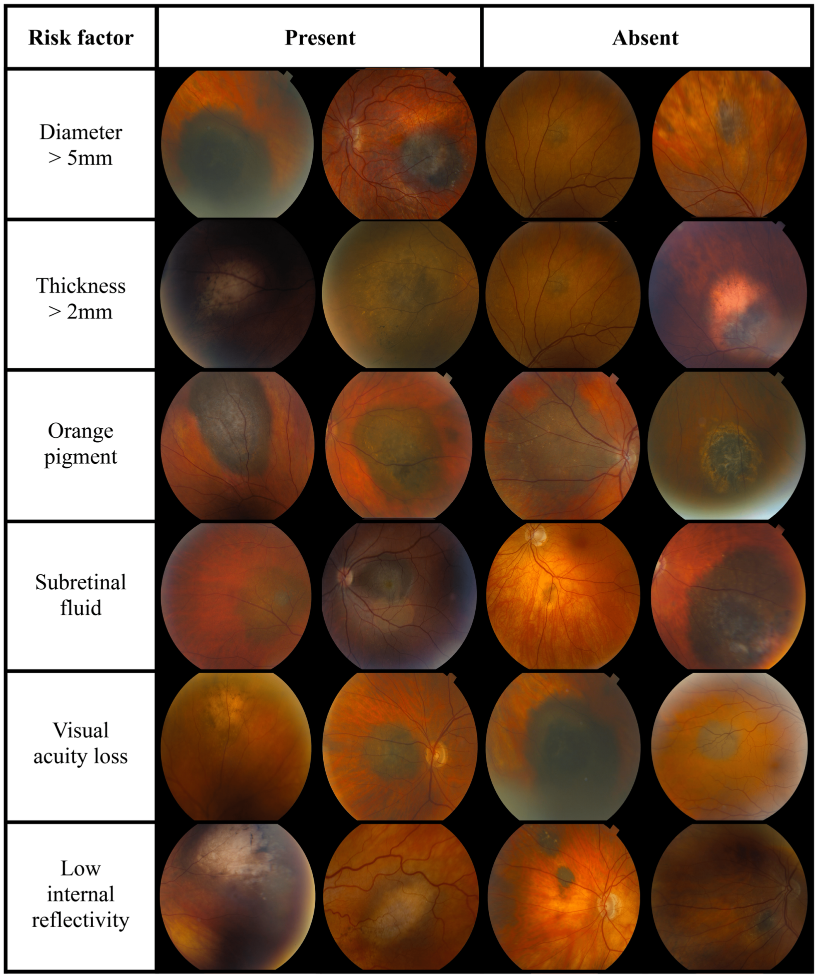

Figure 4.1: Choroidal nevus images with different risk factors

4.2 Control Data

The construction and utilization of control datasets are critical in validating the robustness and specificity of the predictive models developed in this study. Two distinct control sets were prepared, each serving a unique purpose in the experimental design.

4.2.1 Control Set I

Control Set I is built using images extracted from the primary nevus dataset, specifically selected because they do not contain a choroidal nevus. The inclusion of these images in our analyses allows us to examine the model's behavior in the absence of the targeted nevus, while also maximizing the utility of available data. Such a setup is pivotal for assessing whether the model erroneously attributes nevus-associated risk factors to images devoid of nevi. A robust model should ideally demonstrate a low false positive rate in this context, indicating its ability to accurately distinguish between relevant and irrelevant features in the images.

4.2.2 Control Set II

Control Set II was formulated to further test the generalizability and robustness of the model across a broader range of fundus images. This dataset was also procured from Mayo Clinic, Rochester, MN, and consists of 500 images, evenly split between 250 optos and 250 colored fundus images. These images are obtained from patients who do not present any symptoms of choroidal nevus, melanoma, or other ocular conditions with similar clinical symptoms. This set acts as a held-out test set, providing crucial insights into the model's performance on completely independent data. Testing the model on Control Set II helps ascertain its capability to correctly classify images from individuals who are not part of the study's primary patient cohort, thus offering a stringent test of its diagnostic precision in a real-world clinical scenario.

CHAPTER 5

MODEL DESCRIPTIONS

This section describes the AI models employed in this study as the proposed approach or as a baseline for comparison. It outlines their architectural nuances and relevance to the tasks at hand. Each model discussed has been chosen for its unique capabilities in handling specific challenges associated with the analysis of medical imaging data, particularly retinal images. The descriptions provide insights into the foundational principles of each model, their operational mechanisms, and their efficacy in extracting and utilizing visual features crucial for disease diagnosis and prognosis.

5.1 RETFound

Following extensive research on the utility of foundation models, we incorporated RETFound, a foundation model specifically developed for retinal images [65]. The dataset used to train the model contained 1.6 million retinal images collected from various private and public sources. The model was trained in a self-supervised image reconstruction task to learn generalizable representations that are not explicitly labeled yet capture crucial visual features. RETFound has been shown to perform well in downstream tasks across a variety of ocular diseases and imaging conditions, as well as in disease prognosis and predicting systemic diseases. This approach was shown to consistently outperform traditional transfer learning models, offering a more robust framework for tackling complex diagnostic challenges in ophthalmology.

The architecture of RETFound features a large Vision Transformer as the encoder, comprising 24 transformer blocks with an embedding vector size of 1024, while the decoder—used during the self-supervised training phase—consists of a small Vision Transformer with 8 transformer blocks and an embedding vector size of 512 (Figure 5.1). After self-supervised training, the decoder is discarded, and a linear binary head is attached to the encoder for targeted training and inference on downstream tasks. The combination of a pre-trained encoder and a classification head in the model architecture has proven to be an effective low shot learning technique [66]. This is especially relevant for our study, where some risk factors in the dataset exhibit a significant class imbalance, highlighting the utility of this approach in managing sparse and uneven data distributions.

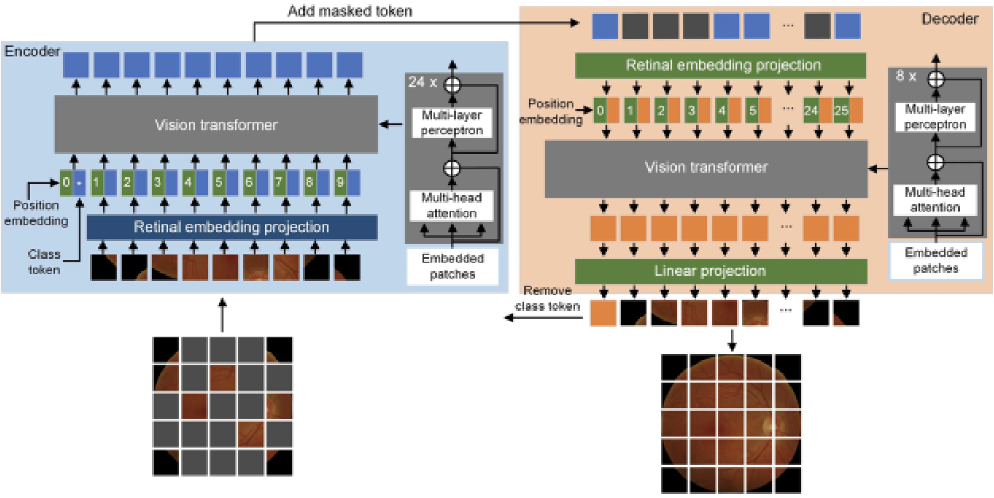

Figure 5.1: RETFound architecture and training procedure

5.2 DINOv2

DINOv2 is an advanced self-supervised learning model that employs Vision Transformers (ViTs) to learn visual representations without requiring labeled data [67]. This model is built on the principle of knowledge distillation, where a student network learns to replicate the output of a simultaneously evolving teacher network. The architecture of DINOv2 typically features multiple transformer blocks, with the teacher network being slightly larger than the student to maintain a robust learning gradient. What makes DINOv2 particularly innovative is its ability to handle complex visual data and extract high-quality features by leveraging the self-distillation approach, effectively teaching itself by refining the student's predictions towards the teacher's outputs. This approach allows DINOv2 to develop a nuanced understanding of visual features that are useful across a broad range of tasks including image classification, object detection, and semantic segmentation. Its ability to operate effectively without labeled data makes it highly popular in scenarios where acquiring annotated datasets is challenging or expensive.

5.3 ResNet

ResNet, short for Residual Network, is a highly influential neural network architecture in the field of deep learning [68]. It introduced the novel concept of residual learning to tackle the degradation problem in training very deep networks. By integrating skip connections, or shortcuts that allow layers to learn residual functions with reference to the layer inputs, ResNet makes it possible to train networks with hundreds, even thousands, of layers effectively. These features address the vanishing gradient problem by allowing the training signal to be directly propagated through the network. ResNet models are popular due to their simplicity and their ability to achieve high accuracy in various visual tasks with less computational complexity compared to deeper, plain networks. The architecture's versatility and efficiency have seen wide adoption in tasks such as image recognition, object detection, and segmentation, making it a widely utilized model in both academic and industrial applications.

CHAPTER 6

IMPLEMENTATION & EXPERIMENTAL DESIGN

This chapter outlines the workflow adopted for processing, analyzing, and validating nevus data within our study. It details the procedures from initial image preprocessing to final model evaluation, highlighting the methodologies employed to ensure robust and generalizable results in clinical diagnostics.

6.1 Nevus Data Preprocessing

For the study, the original fundus images, captured at a resolution of 2392×2048 pixels, were carefully processed to ensure quality and consistency. For each patient we chose a single comprehensive image that captured the entire nevus, and if additional images exhibited varying angles of the same nevus, these too were incorporated into the analysis. Images compromised by significant noise, motion blur, or lighting artifacts were excluded from the study. Furthermore, any cases lacking clear, complete views of the colored fundus in a single frame were omitted. Ultra-widefield (UWF) images were also excluded from the study due to the added challenges that their small number and restricted visibility presented. After narrowing down the collection, the study proceeded with 1540 images from 384 patients, an average of 4.0 ± 2.3 images per patient. Each selected image was resized to 224 x 224 pixels to comply with the proposed architecture's input requirements. To enhance data variability and robustness against overfitting, 2 random augmentations were applied from a set of 15 possibilities using the RandAugment function from the PyTorch Image Models (*timm*) library [69]. This preprocessing approach is utilized to reduce overfitting and enhance the AI model's ability to reliably interpret and learn from visual inputs.

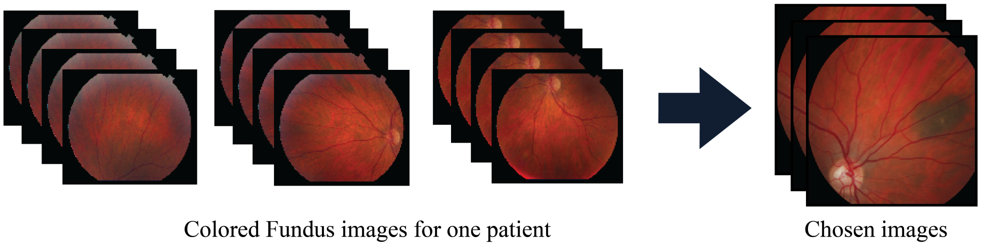

Figure 6.1: Nevus data preparation

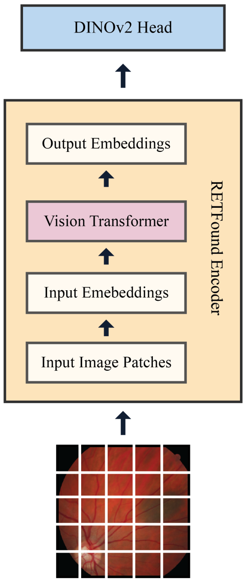

Figure 6.2: Proposed model — RETFound + DINOv2 Head

6.2 Experimental Setup

We employed stratified k-fold cross-validation to address the challenges posed by an imbalanced dataset and a limited number of images for training and validation. Data was divided into five distinct folds; in each cross-validation cycle, four folds were designated for training and validation, while the fifth served as the test set. This strategy enabled us to leverage the entire dataset for testing by rotating the test fold, thereby enhancing the reliability of our performance metrics. The Area Under the Receiver Operating Characteristic curve (AUC) was computed across all images to evaluate the average performance of the model, ensuring that our results reflected the model's capability across the entire dataset rather than a restricted subset.

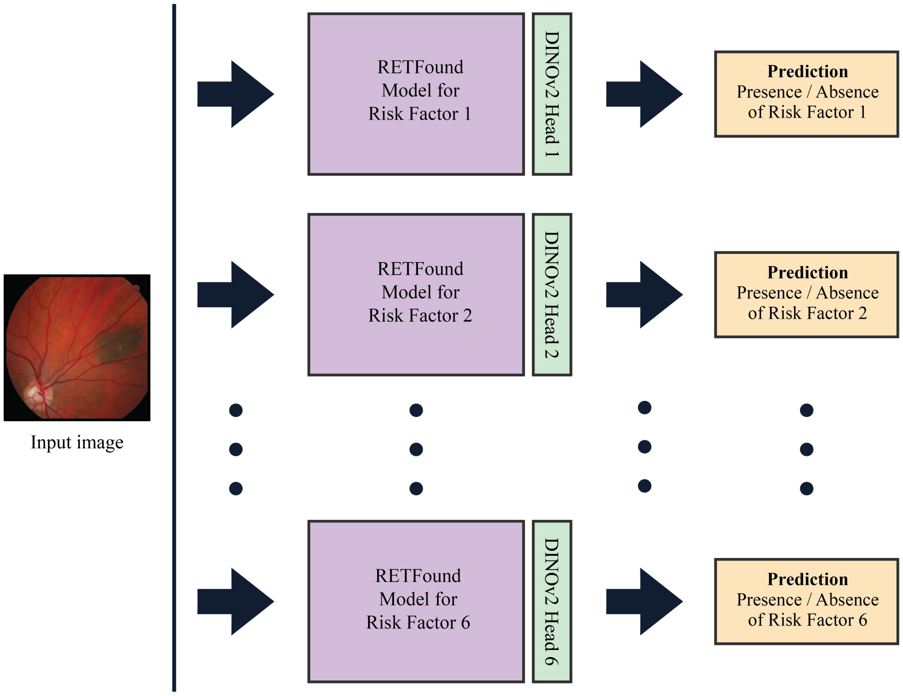

Figure 6.3: Model training/validation workflow

To combat the challenge of a dataset imbalance, we oversampled images from the minority class during training. This oversampling approach may introduce overfitting, however the data augmentations introduced (Section 6.1) mitigate this potential issue. In parallel, we also ran experiments utilizing weights based on class ratio rather than oversampling and compared the results. Both strategies employed specific loss functions to optimize our models effectively. For the oversampling strategy, the standard cross entropy loss function was used, defined as:

$$L = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{6.1}$$

Where L represents the loss, M is the number of classes, $y_{o,c}$ is a binary indicator for whether class c is the correct classification for observation o, and $p_{o,c}$ is the predicted probability for observation o belonging to class c.

For the weighted strategy, the weighted cross entropy loss was utilized, described by:

$$L = -\sum_{c=1}^{M} w_c \cdot y_{o,c} \log(p_{o,c}) \tag{6.2}$$

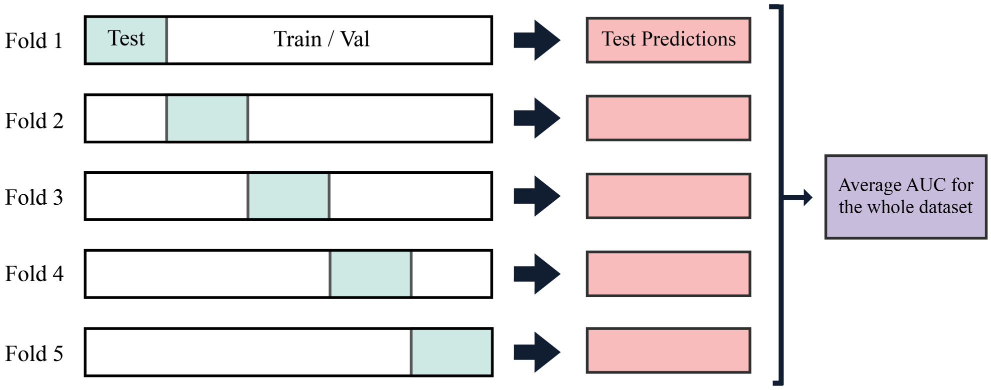

Figure 6.4: Model training strategy

In this equation, w_c are the weights applied to each class to account for imbalances, with other terms defined as in the standard cross entropy.

We trained six different binary classification models targeting each identified risk factor respectively (Figure 6.3). Models were trained using the AdamW optimizer with a batch size of 16 and a base learning rate of $5e^{-3}$. Each model underwent a training process spanning 300 epochs to achieve convergence. Models with the best validation loss were chosen for testing purposes (Figure 6.5). Training was performed on Mayo Clinic's HPC cluster, mForge, which contains A100 and V100 GPUs.

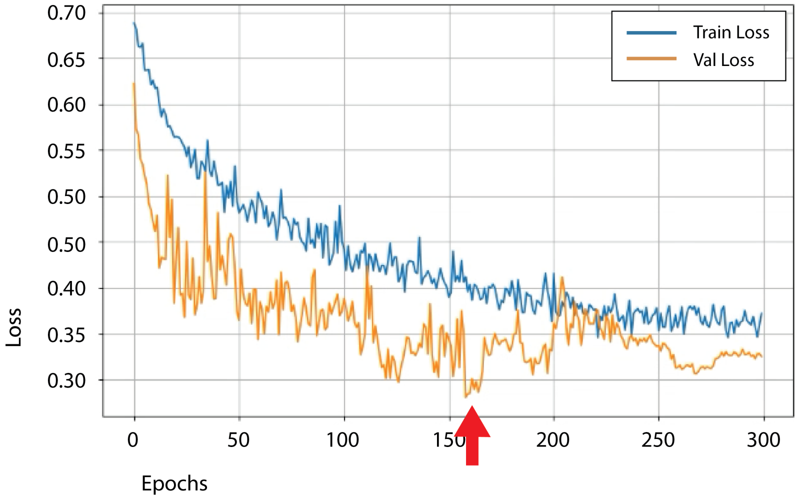

Figure 6.5: Model choice based on best validation loss (red arrow)

6.3 Controls Setup

Control Set I was screened to eliminate images affected by motion blur or adverse lighting, similar to the primary nevus dataset. This subset contained 4,195 images, averaging 9.2 ± 6.0 images per patient. From Control Set II, we excluded Ultra-widefield (UWF) images, consistent with the primary nevus dataset's criteria, resulting in a subset of 250 images. After tuning hyperparameters with the validation set and reporting the AUC on the test data, the entire nevus dataset was used to train the models. These models were then evaluated on the control sets. We did not use AUC as the metric for the control test because the ROC curve requires both positive and negative instances to plot true positive rate against false positive rate. With only one class, calculating these rates is impossible. Therefore, we chose accuracy to evaluate the AI models on the control sets.

6.4 Nevus Detector Setup

To evaluate the robustness and adaptability of the proposed architecture for different tasks, we conducted an additional test to predict the presence of a choroidal nevus in an image. We used the primary nevus dataset for images containing nevi and Control Set II for images without nevi. The model was trained and validated using the same strategies as the risk factor classification models, employing both oversampling and weighted loss techniques.

CHAPTER 7

RESULTS

This chapter presents the results of our analysis using the datasets and experimental setup described in the previous sections. The findings showcase high performance across various diagnostic models, providing detailed insights into patient attributes and associated risk factors. We also visualize various model performance metrics for a comparative analysis among different risk factors and datasets.

7.1 Patient Data Analysis

Patients in the study that were preprocessed for experimentation were primarily White (99.0%) followed by Hispanic (0.5%) and Asian and Middle Eastern (0.25% each). The distribution of clinical features and risk factors is shown in Table 7.1. 61.5% of the patients in the study are females, compared to 38.5% males. The average age of the patients is 61.8 years old. Among the risk factors, the *diameter* $\geq 5mm$ risk factor is most widely present (48.3%), followed by the *low internal reflectivity* risk factor (31.7%) and the *orange pigment* risk factor (14.6%).

Interestingly, and crucially for this study, we found that 161 (39.9%) affected eyes in the preprocessed dataset contain a nevus but do not exhibit any of the risk factors (Table 7.2). Only 8 (2.0%) affected eyes contain 4 risk factors, while none of the affected eyes contain either 5 or 6 risk factors. The significant fraction of patients not exhibiting any risk factors highlights the potential of our AI model to reduce unnecessary referrals, thereby reducing the burden on healthcare systems and escalate care for patients who truly require specialist intervention.

Table 7.1: Summary of patient attributes and associated risk factors

Attribute	Quantity	Risk Factor Y/N (%)	Total
Sex (M/F)	148 (38.5%) / 236 (61.5%)	-	384
Age (mean ± std)	61.8 ± 17.0	-	-
Distance to ON (mm)	4.8 ± 3.7	-	-
Distance to fovea (mm)	4.4 ± 3.9	-	-
Diameter (mm)	5.7 ± 3.1	195 / 209 (48.3 / 51.7)	404
Thickness (mm)	1.0 ± 0.7	23 / 381 (5.7 / 94.3)	
Orange Pigment	-	59 / 345 (14.6 / 85.4)	
Subretinal Fluid	-	13 / 391 (3.2 / 96.8)	
Visual Acuity	-	22 / 382 (5.4 / 94.6)	
Internal Reflectivity	-	128 / 276 (31.7 / 68.3)	

Table 7.2: Number of risk factors per eye in dataset

Risk Factors	Number of Eyes
0	161 (39.9%)
1	103 (25.5%)
2	92 (22.5%)
3	41 (10.1%)
4	8 (2.0%)
5	0 (0.0%)
6	0 (0.0%)
Total	404 (100%)

7.2 Risk Factor Classification

The effectiveness of the proposed diagnostic approach is illustrated in Table 7.3. The *Diameter* $\geq 5mm$ model achieved the highest performance, with an AUC of 0.93, followed by the *Thickness* $\geq 2mm$ and *Low internal reflectivity* models with AUC of 0.84 and 0.82, respectively. Figure 7.1 presents the ROC plots for all risk factors stacked up against each other, and Figure 7.2 summarizes the average AUC on the train, validation, and test sets for the model with the best validation AUC for each of the risk factors. The models were thoroughly tested in Chapter 8 to validate their robustness and interpretability.

Table 7.3: Risk factor classification results

Risk Factor	AUC
Diameter $\geq$ 5mm	0.93 ± 0.05
Thickness $\geq$ 2mm	0.84 ± 0.07
Orange pigment	0.70 ± 0.07
Subretinal fluid	0.77 ± 0.08
Visual acuity loss	0.63 ± 0.10
Low int. reflectivity	0.82 ± 0.06

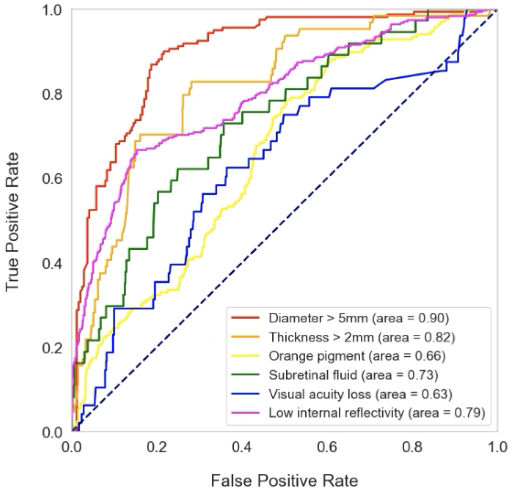

Figure 7.1: ROC curves for risk factor classification

7.3 Controls Data Evaluation

Tables 7.4 and 7.5 summarize the evaluation results for the AI models on Control Sets I and II, respectively. The models achieved an average accuracy of 97.9% for Control Set I and 99.3% for Control Set II across all risk factors. This high accuracy indicates that the models effectively distinguish normal images from those containing nevi or those at risk of developing into melanoma, minimizing the likelihood of false positives in clinical assessments.

7.4 Summarizing Model Performance for All Datasets

Figure 7.3 summarizes the distribution of the model predicting the presence of a risk factor when the ground truth was the presence/absence of the risk factor, and for the two control sets. For all risk factors, the model was more

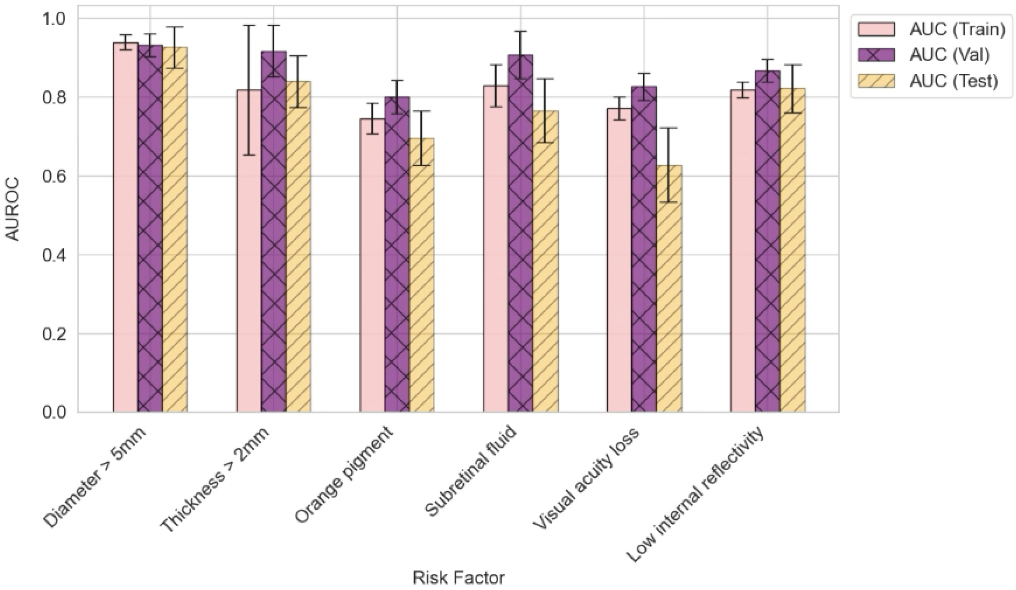

Figure 7.2: AUC for train, val, and test set for risk factor classification

confident predicting a risk factor when it was actually there, compared to when it was now. For the controls data, especially, the models were very confident in their prediction as the violin plots are highly skewed towards a lower confidence. The violin and swarm plot distributions highly correlate with the AUC, as expected; the higher the AUC, the more the distributions for the presence/absence of nevus are separated from one another.

Table 7.4: Evaluating proposed model on Control Set I

Risk Factor	Accuracy
Diameter $\geq$ 5mm	0.98
Thickness $\geq$ 2mm	1.00
Orange pigment	0.94
Subretinal fluid	1.00
Visual acuity loss	1.00
Low int. reflectivity	0.97

7.5 Nevus Detection

Table 7.6 presents the results of nevus detection using the two proposed strategies: weighted loss function and oversampling. The models achieved an

Table 7.5: Evaluating proposed model on Control Set II

Risk Factor	Accuracy
Diameter $\geq$ 5mm	1.00
Thickness $\geq$ 2mm	1.00
Orange pigment	0.97
Subretinal fluid	1.00
Visual acuity loss	1.00
Low int. reflectivity	1.00

AUC of 0.89 ± 0.11 with the weighted loss function strategy, and an improved AUC of 0.91 ± 0.05 with the oversampling strategy. These results indicate that oversampling provides a more robust performance for nevus detection, leading to more reliable identification of nevi and potentially reducing the risk of missing early signs of melanoma in clinical settings.

Table 7.6: Nevus detection results

Strategy	AUC
Weighted loss function	0.89 ± 0.11
Oversampling	0.91 ± 0.05

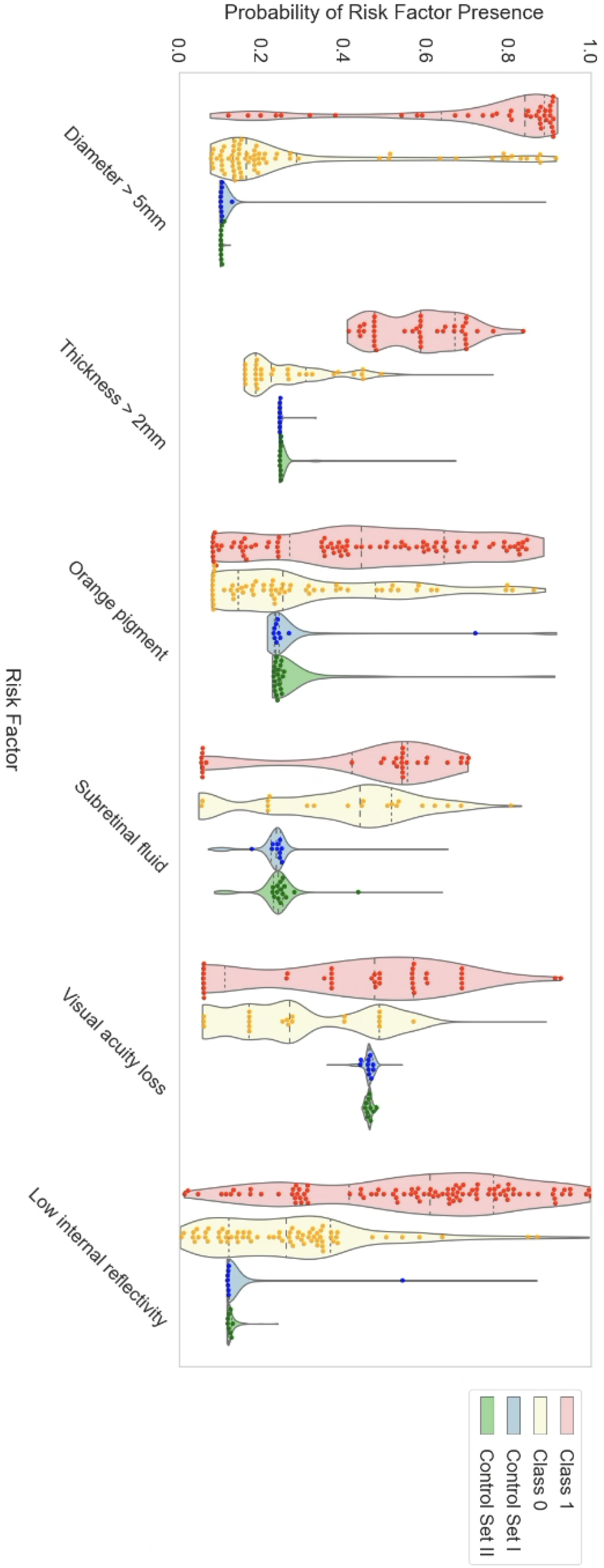

Figure 7.3: Distribution of risk factor prediction for different data categories

CHAPTER 8

DISCUSSION & FUTURE WORK

This section delves into the critical aspects of model evaluation, particularly focusing on our architecture's comparative performance against established baselines. We discuss the model's interpretability and the techniques employed to ensure it is understandable and accessible to both medical professionals and patients, thereby increasing confidence in its use. This chapter also explores the implications of the findings for future research and potential enhancements in model design and deployment.

8.1 Model Performance Comparison

Table 8.1 summarizes a comparative analysis between the proposed architecture and various configurations of ResNet. These results illustrate that our model consistently outperforms ResNet across various test scenarios, even with smaller input image sizes, thus indicating our model's robustness and its refined capability to analyze retinal images efficiently.

Table 8.1: Comparison of AUC scores across different models and resolutions

Risk Factor	Ours (224x224)	ResNet-18 (224x224)	ResNet-50 (224x224)	ResNet-18 (512x512)
Diameter $\geq$ 5mm	**0.93 ± 0.05**	0.92 ± 0.02	0.91 ± 0.02	0.87 ± 0.03
Thickness $\geq$ 2mm	**0.84 ± 0.07**	0.77 ± 0.04	0.80 ± 0.07	0.54 ± 0.13
Orange pigment	**0.70 ± 0.07**	0.62 ± 0.03	0.64 ± 0.07	0.48 ± 0.07
Subretinal fluid	**0.77 ± 0.08**	0.44 ± 0.28	0.73 ± 0.20	0.45 ± 0.10
Visual acuity loss	**0.63 ± 0.10**	0.53 ± 0.21	0.53 ± 0.06	0.48 ± 0.15
Low int. feflectivity	**0.82 ± 0.06**	0.80 ± 0.06	0.76 ± 0.09	0.80 ± 0.07

8.2 Model Interpretation

8.2.1 Understanding Model Behaviour For Varying Risk Factors

To validate our model's effectiveness, we utilized t-SNE to visualize high-dimensional embeddings in two dimensions. Figure 8.1 illustrates this, with data point colors indicating actual nevus diameters and thickness, respectively. While t-SNE is not able to generate distinct clustering between data points with and without the risk factor, the visualization approximates a separation of nevi diameters around the 5mm threshold and nevus thickness around the 2mm threshold even for highly compressed data points.

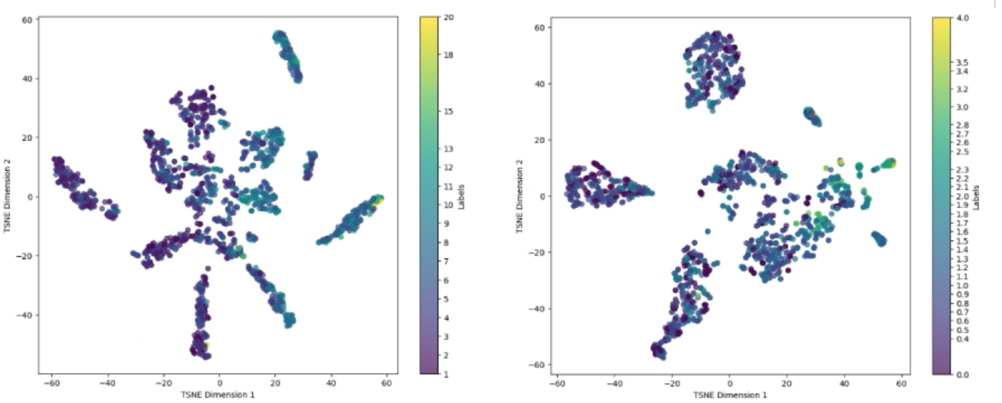

Figure 8.1: t-SNE visualization for nevus diameter (left) and thickness (right)

Furthermore, we illustrate model output confidence against the actual diameters and thickness of choroidal nevi in the retinal images (Figure 8.2). We show that the model's confidence in the presence or absence of the *Diameter ¿ 5mm* risk factor increases as the nevus diameter increases or decreases, respectively. We observe a similar trend for the *Thickness ¿ 2mm* risk factor as well, although the separation is less pronounced compared to the diameters. These experiments shed further light on the AI models' interpretability.

8.2.2 Saliency Maps

We generated saliency maps using Grad-CAM to provide insights into how the model focuses on specific image areas to determine risk factors (Figure

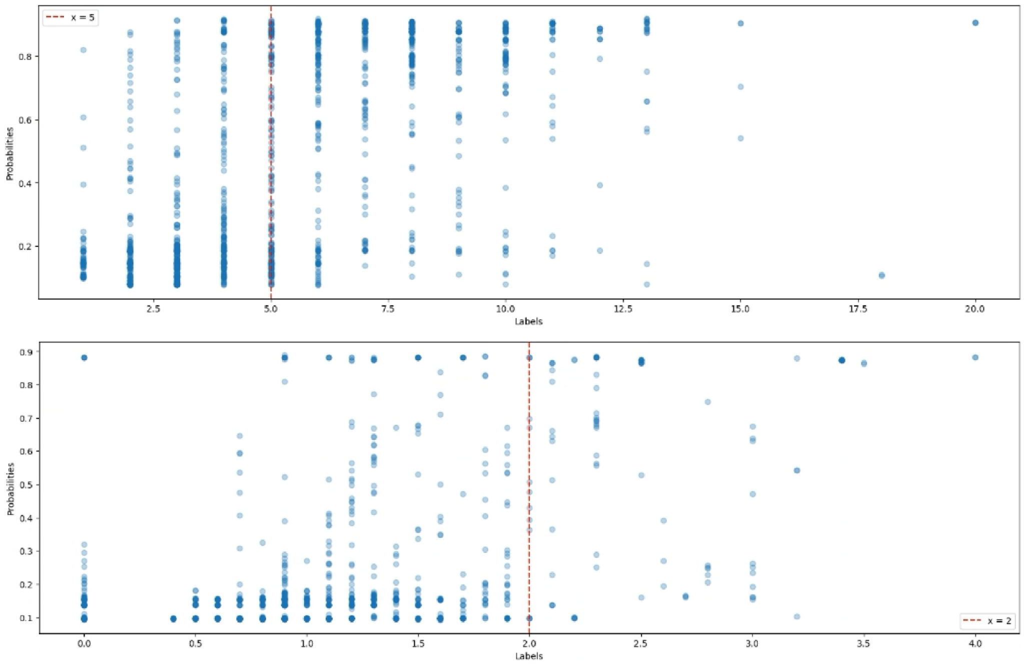

Figure 8.2: Model confidence for varying nevus diameter (top) and thickness (bottom)

8.3). While Grad-CAM is conventionally utilized for convolutional neural networks (CNNs), we apply it to the final output of the encoder, which has the dimensions $N \times 197 \times 1024$, where N represents the batch size. In the dimension with 197 elements, the first element is designated as the class token. Consequently, the remainder of the output is interpreted as a 14×14 spatial image with 1024 channels, upon which Grad-CAM is employed. These maps reveal that the model primarily concentrates on the nevus and its boundaries, highlighting the anatomical features most relevant for risk assessment.

8.2.3 Evaluation of Diameter for Varying Zoom Levels

Recognizing the potential deployment across varied imaging technologies, we conducted experiments to assess model performance under different camera zoom levels. By simulating various zoom scenarios through image cropping and padding, we determined the model's robustness across a spectrum of visual scales. Results are shown in Table 8.2; the maximum drop in performance observed is only 5.28% which confirms the model's consistent per-

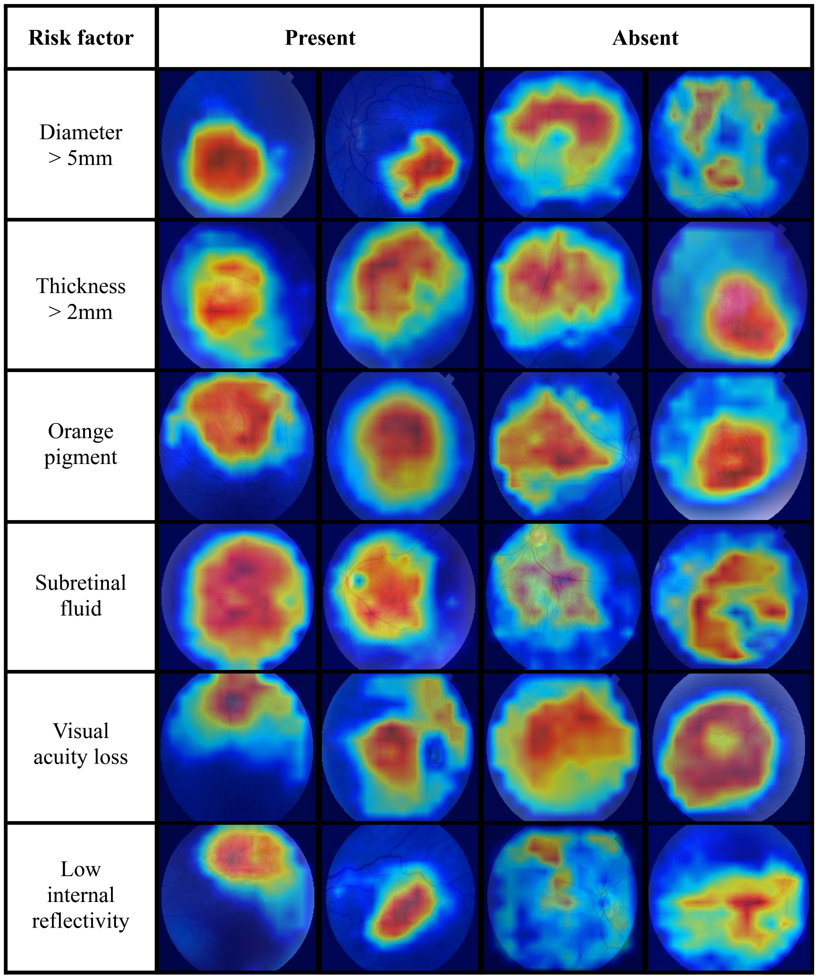

Figure 8.3: Saliency maps for different risk factors

formance, ensuring reliability regardless of the imaging equipment used in clinical practice.

8.3 Limitations

This study faces several limitations that must be acknowledged. Firstly, the dataset comprises images from a single institution, which may not represent the demographic and clinical diversity seen across different geographic locations. This limitation could affect the generalizability of the findings to other populations. Additionally, certain risk factors are underrepresented in our dataset, such as the presence of subretinal fluid, which is observed in only 3.2% of the patients that are part of this study. Despite employing advanced

Table 8.2: AUC for different zoom levels - Diameter $\geq$ 5mm

Zoom in/out	Zoom level	AUC	Percentage drop
-	-	0.93 ± 0.05	-
In	20%	0.89 ± 0.02	3.78%
	40%	0.88 ± 0.01	4.64%
Out	20%	0.88 ± 0.01	4.64%
	40%	0.88 ± 0.02	5.28%

data augmentation techniques to address this issue, the scarcity of examples limits our ability to thoroughly evaluate the model's performance for these specific risk factors. A more robust validation could be achieved with an independent test set that includes a wider variety of scenarios featuring these underrepresented risk factors, further strengthening the study's conclusions and applicative reliability.

8.4 Future Work

Future research could focus on several key areas to extend the findings of this study:

- Integration with other imaging modalities: While this study primarily utilized choroidal nevi retinal images, future studies could incorporate the ability to add other imaging modalities if available, such as optical coherence tomography (OCT) and ultrasonography, to enhance diagnostic accuracy and validate the findings.
- Expansion of the dataset: To improve the generalizability of the AI models, future work could include a more diverse dataset that covers different demographic groups and geographic locations. It could also include data that exhibits certain risk factors, such as subretinal fluid, even in the absence of choroidal nevi; the goal would then be to learn representations of particular risk factors from a wide variety of images, thereby improving generalizability.
- Real-time diagnostic tools: Development of a real-time diagnostic tool that can be integrated into clinical workflows, providing instant risk factor assessments for choroidal nevi transformation into melanoma.

- Longitudinal studies: Conducting longitudinal studies to track the progression of choroidal nevi over time in the same patients, which could provide more insights into the natural history and risk factors associated with their transformation into melanoma.

- Explainability and user trust: Further exploration into model explainability to increase trust among clinicians by providing understandable insights into how the AI models make their predictions.

CHAPTER 9

CONCLUSION

This study has demonstrated the potential of artificial intelligence in enhancing the detection and characterization of choroidal nevi and their transformation into melanoma. By leveraging advanced machine learning techniques and a substantial dataset of retinal images, the developed models show promising results in identifying critical risk factors associated with the disease. The findings not only contribute to the scientific community by improving understanding of choroidal nevi but also pave the way for future innovations in ophthalmic imaging and diagnostics. The implications of this research are profound, offering a pathway toward more accurate, efficient, and accessible diagnostics in ophthalmology.

REFERENCES

[1] M. D. C. Terveen, “Ophthalmology numbers cause for concern,” Aug 2022.

[2] Y. Xu, L. Lou, Y. Wang, Q. Miao, K. Jin, M. Chen, and J. Ye, “Epidemiological study of uveal melanoma from us surveillance, epidemiology, and end results program (2010-2015),” *Journal of ophthalmology*, vol. 2020, p. 3614039, Feb 2020.

[3] S. Kaliki, C. L. Shields, and J. A. Shields, “Uveal melanoma: estimating prognosis,” *Indian journal of ophthalmology*, vol. 63, no. 2, p. 93–102, Feb 2015.

[4] E. Kujala, T. Mäkitie, and T. Kivelä, “Very long-term prognosis of patients with malignant uveal melanoma,” *Investigative ophthalmology visual science*, vol. 44, no. 11, p. 4651–9, Nov 2003.

[5] J. L. Chien, K. Sioufi, T. Surakiatchanukul, J. A. Shields, and C. L. Shields, “Choroidal nevus: a review of prevalence, features, genetics, risks, and outcomes,” *Current opinion in ophthalmology*, vol. 28, no. 3, p. 228–237, May 2017.

[6] M. B. Qureshi, P. C. Lentz, T. T. Xu, L. J. White, T. W. Olsen, J. S. Pulido, and L. A. Dalvin, “Choroidal nevus features associated with subspecialty referral,” *Ophthalmology. Retina*, vol. 7, no. 12, p. 1097–1108, Dec 2023.

[7] A. Stuart, “Neuro: How to minimize diagnostic errors,” Sep 2020. [Online]. Available: https://www.aao.org/eyenet/article/neuro-how-to-minimize-diagnostic-errors

[8] A. Panesar, *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes.* Apress, Feb 2019.

[9] A. Al Kuwaiti, K. Nazer, A. Al-Reedy, S. Al-Shehri, A. Al-Muhanna, A. V. Subbarayalu, D. Al Muhanna, and F. A. Al-Muhanna, “A review of the role of artificial intelligence in healthcare,” *Journal of Personalized Medicine*, vol. 13, no. 6, p. 951, Jun 2023.

[10] A. Nucci, "Large language models in healthcare: Use cases and benefits," Dec 2023. [Online]. Available: https://aisera.com/blog/large-language-models-healthcare/

[11] S. Daley, "Ai in healthcare: Uses, examples and benefits," Jun 2022. [Online]. Available: https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare

[12] C. Mishra and K. Tripathy, "Fundus camera," In: StatPearls [Internet], Treasure Island, FL, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK585111/

[13] M. M. T. Witmer and S. Kiss, "The clinical utility of ultra-wide-field imaging," Mar 2012.

[14] S. Allarakha, "What is the difference between a-scan and b-scan?" Jan 2021.

[15] R. G. Waldron, "A-scan biometry: Practice essentials, ultrasound principles, ultrasound biometry instrumentation." [Online]. Available: https://emedicine.medscape.com/article/1228447-overview?form=fpf

[16] D. Urban, N. Lighthizer, J. Ramey, and R. Bunch, "Scoring an a+ on a b-scan," *Review of Optometry*, Mar 2016. [Online]. Available: https://www.reviewofoptometry.com/article/scoring-an-a-on-a-bscan

[17] [Online]. Available: https://www.bumrungrad.com/en/treatments/fundus-fluorescein-angiography-ffa-indocyanine-green-icg-angiography

[18] [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/24618-fluorescein-angiography

[19] [Online]. Available: https://medicine.uiowa.edu/eye/patient-care/imaging-services/indocyanine-green-angiography

[20] C. C. m. professional, "Optical coherence tomography." [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/17293-optical-coherence-tomography

[21] C. Medina and A. Singh, "Imaging of intraocular tumors," Oct 2014.

[22] [Online]. Available: https://eyewiki.aao.org/Fundus_Autofluorescence

[23] K. H. Nguyen, B. C. Patel, and P. Tadi, "Anatomy, head and neck: Eye retina," *StatPearls*, Aug 2023.

[24] [Online]. Available: https://retina-international.org/retinal-health-2/our-eyes-new/

[25] J. Morales, "Assessment of iris reflection artifacts and alignment in fundus images," Ph.D. dissertation, 07 2011.

[26] C. L. Shields, J. Cater, J. A. Shields, A. D. Singh, M. C. Santos, and C. Carvalho, "Combination of clinical factors predictive of growth of small choroidal melanocytic tumors," *Archives of ophthalmology (Chicago, Ill.: 1960)*, vol. 118, no. 3, p. 360–4, Mar 2000.

[27] C. L. Shields, L. A. Dalvin, D. Ancona-Lezama, M. D. Yu, M. Di Nicola, J. Williams, Basil K, J. A. Lucio-Alvarez, S. M. Ang, S. Maloney, R. J. Welch, and J. A. Shields, "Choroidal nevus imaging features in 3,806 cases and risk factors for transformation into melanoma in 2,355 cases: The 2020 taylor r. smith and victor t. curtin lecture," *Retina (Philadelphia, Pa.)*, vol. 39, no. 10, p. 1840–1851, Oct 2019.

[28] A. Cheung, I. Scott, T. Murray, and C. Shields, "Distinguishing a choroidal nevus from a choroidal melanoma," Feb 2012. [Online]. Available: https://www.aao.org/eyenet/article/distinguishing-choroidal-nevus-from-choroidal-mela

[29] M. Schlitt, S. Kaliki, and C. Shields, "Reduction of orange pigment overlying choroidal melanoma following plaque radiotherapy," Sep 2013.

[30] B. Damato, "Pigmented fundus lesions." [Online]. Available: https://www.college-optometrists.org/clinical-guidance/clinical-management-guidelines/pigmentedfunduslesions

[31] Y.-K. Kim, S. J. Woo, K. E. Kim, and K. H. Park, "Subretinal fluid resorption," *Ophthalmology*, vol. 117, no. 8, pp. 1655–1655.e4, Aug 2010.

[32] [Online]. Available: https://www.aoa.org/healthy-eyes/vision-and-vision-correction/visual-acuity?sso=y

[33] C. to Wikimedia projects, "Snellen chart," Apr 2024. [Online]. Available: https://en.wikipedia.org/wiki/Snellen_chart

[34] H. Sánchez-Tocino, A. García-Layana, A. Salinas-Alamán, J. M. Alcalde-Navarrete, A. Panizo-Santos, and R. Martínez-Monge, "Central retinal vascular occlusion by orbital pseudotumor," *Retina*, vol. 24, no. 3, p. 455–458, Jun 2004.

[35] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press, May 2014.

[36] R. Pramoditha, "Overview of a neural network's learning process - data science 365 - medium," *Data Science 365*, Feb 2022. [Online]. Available: https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[38] A. Kolesnikov, A. Dosovitskiy, D. Weissenborn, G. Heigold, J. Uszkoreit, L. Beyer, M. Minderer, M. Dehghani, N. Houlsby, S. Gelly, T. Unterthiner, and X. Zhai, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.

[39] “What are foundation models for ai?_2023,” Sep 2023. [Online]. Available: https://www.redhat.com/en/topics/ai/what-are-foundation-models

[40] T. Fawcett, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016786550500303X

[41] N. A. Ahmed, “What is a confusion matrix in machine learning? the model evaluation tool explained,” *DataCamp*, Nov 2023. [Online]. Available: https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning

[42] C. to Wikimedia projects, “Receiver operating characteristic,” Jun 2024.

[43] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[44] N. Shukla, “Gradient weighted class activation map(grad-cam) - ninad shukla,” *Medium*, Jul 2019. [Online]. Available: https://medium.com/@ninads79shukla/gradcam-73a752d368be

[45] C. to Wikimedia projects, “T-distributed stochastic neighbor embedding,” Feb 2024. [Online]. Available: https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

[46] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, and J. S. Utikal, “Deep neural networks are superior to dermatologists in melanoma image classification,” *European Journal of Cancer*, vol. 119, pp. 11–17, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959804919303491

[47] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Opthalmology amp; Visual Science*, vol. 57, no. 13, p. 5200, Oct 2016.

[48] V. Gulshan, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal," *JAMA*, vol. 316, no. 22, p. 2402–2410, Dec 2016.

[49] P. Heydon, C. Egan, L. Bolter, R. Chambers, J. Anderson, S. Aldington, I. M. Stratton, P. H. Scanlon, L. Webster, S. Mann, A. d. Chemin, C. G. Owen, A. Tufail, and A. R. Rudnicka, "Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients," *British Journal of Ophthalmology*, vol. 105, no. 5, p. 723–728, May 2021.

[50] T. Nagasawa, H. Tabuchi, H. Masumoto, H. Enno, M. Niki, Z. Ohara, Y. Yoshizumi, H. Ohsugi, and Y. Mitamura, "Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy," *International Ophthalmology*, vol. 39, no. 10, p. 2153–2159, Feb 2019.

[51] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration oct images," *Ophthalmology Retina*, vol. 1, no. 4, p. 322–327, Jul 2017.

[52] S. Natarajan, "Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence," *JAMA Ophthalmology*, vol. 137, no. 10, p. 1182–1188, Oct 2019.

[53] S. Chikumba, Y. Hu, and J. Luo, "Deep learning-based fundus image analysis for cardiovascular disease: a review," *Therapeutic Advances in Chronic Disease*, vol. 14, p. 20406223231209895, 2023. [Online]. Available: https://doi.org/10.1177/20406223231209895

[54] K. M. Kim, T.-Y. Heo, A. Kim, J. Kim, K. J. Han, J. Yun, and J. K. Min, "Development of a fundus image-based deep learning diagnostic tool for various retinal diseases," *Journal of Personalized Medicine*, vol. 11, no. 5, Apr 2021.

[55] J. Wu, R. Hu, Z. Xiao, J. Chen, and J. Liu, "Vision transformer-based recognition of diabetic retinopathy grade," *Medical Physics*, vol. 48, no. 12, p. 7850–7863, Nov 2021.

[56] K. Xu, S. Huang, Z. Yang, Y. Zhang, Y. Fang, G. Zheng, B. Lin, M. Zhou, and J. Sun, "Automatic detection and differential diagnosis of age-related macular degeneration from color fundus photographs using deep learning with hierarchical vision transformer," *Computers in Biology and Medicine*, vol. 167, p. 107616, Dec 2023.

[57] H. Zhang, Y. Liu, K. Zhang, S. Hui, Y. Feng, J. Luo, Y. Li, and W. Wei, "Validation of the relationship between iris color and uveal melanoma using artificial intelligence with multiple paths in a large chinese population," *Frontiers in Cell and Developmental Biology*, vol. 9, Aug 2021.

[58] T. K. Yoo, J. Y. Choi, H. K. Kim, I. H. Ryu, and J. K. Kim, "Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images," *Computer Methods and Programs in Biomedicine*, vol. 205, p. 106086, Jun 2021.

[59] E. C. Zabor, V. Raval, S. Luo, D. E. Pelayes, and A. D. Singh, "A prediction model to discriminate small choroidal melanoma from choroidal nevus," *Ocular Oncology and Pathology*, vol. 8, no. 1, p. 71–78, Dec 2021.

[60] A. K. Dadzie, S. Iddir, M. Abtahi, B. Ebrahimi, D. Le, T. Son, M. J. Heiferman, and X. Yao, "Deep learning for automated diagnosis of uveal melanoma," in *Ophthalmic Technologies XXXIV*. SPIE, Mar 2024. [Online]. Available: http://dx.doi.org/10.1117/12.3000655

[61] C. L. Shields, S. E. Lally, L. A. Dalvin, M. S. Sagoo, M. Pellegrini, S. Kaliki, A. K. Gündüz, M. Furuta, P. Mruthyunjaya, A. T. Fung, J. S. Duker, S. M. Selig, A. Yaghy, S. R. Ferenczy, M. B. Eydelman, and M. S. Blumenkranz, "White paper on ophthalmic imaging for choroidal nevus identification and transformation into melanoma," *Translational Vision Science Technology*, vol. 10, no. 2, p. 24–24, Feb 2021.

[62] S. Iddir, J. Love, J. Ma, J. Bryan, S. Ganesh, M. Heiferman, and D. Yi, "Predicting malignant transformation of choroidal nevi using machine learning," Dec 2023. [Online]. Available: https://www.researchsquare.com/article/rs-3778562/v1

[63] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomedical Engineering*, vol. 2, no. 3, p. 158–164, Feb 2018.

[64] S. Kaliki, V. Vempuluru, N. Ghose, G. Patil, R. Viriyala, and K. Dhara, "Artificial intelligence and machine learning in ocular oncology: Retinoblastoma," *Indian Journal of Ophthalmology*, vol. 71, no. 2, p. 424, 2023.

[65] Y. Zhou, M. A. Chia, S. K. Wagner, M. S. Ayhan, D. J. Williamson, R. R. Struyven, T. Liu, M. Xu, M. G. Lozano, P. Woodward-Court, Y. Kihara, A. Altmann, A. Y. Lee, E. J. Topol, A. K. Denniston, D. C. Alexander, and P. A. Keane, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, p. 156–163, Sep 2023.

[66] P. Burlina, W. Paul, P. Mathew, N. Joshi, K. D. Pacheco, and N. M. Bressler, "Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases," *JAMA ophthalmology*, vol. 138, no. 10, p. 1070–1077, Oct 2020.

[67] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "Dinov2: Learning robust visual features without supervision," Apr 2023.

[68] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Dec 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[69] [Online]. Available: https://timm.fast.ai/